INTERMITTENT DEMAND FORECASTING WITH INTEGER AUTOREGRESSIVE MOVING AVERAGE MODELS

A Thesis submitted for the degree of Doctor of Philosophy

By

Maryam Mohammadipour

School of Business and Management

Faculty of Enterprise and Innovation

Buckinghamshire New University

Brunel University

April, 2009

Abstract

This PhD thesis focuses on using time series models for counts in modelling and forecasting a special type of count series called intermittent series. An intermittent series is a series of non-negative integer values with some zero values. Such series occur in many areas including inventory control of spare parts. Various methods have been developed for intermittent demand forecasting with Croston's method being the most widely used.

Some studies focus on finding a model underlying Croston's method. With none of these studies being successful in demonstrating an underlying model for which Croston's method is optimal, the focus should now shift towards stationary models for intermittent demand forecasting.

This thesis explores the application of a class of models for count data called the Integer Autoregressive Moving Average (INARMA) models. INARMA models have had applications in different areas such as medical science and economics, but this is the first attempt to use such a model-based method to forecast intermittent demand.

In this PhD research, we first fill some gaps in the INARMA literature by finding the unconditional variance and the autocorrelation function of the general INARMA(p,q) model. The conditional expected value of the aggregated process over lead time is also obtained to be used as a lead time forecast. The accuracy of h-step-ahead and lead time INARMA forecasts are then compared to those obtained by benchmark methods of Croston, Syntetos-Boylan Approximation (SBA) and Shale-Boylan-Johnston (SBJ).

The results of the simulation suggest that in the presence of a high autocorrelation in data, INARMA yields much more accurate one-step ahead forecasts than benchmark methods. The degree of improvement increases for longer data histories. It has been shown that instead of identification of the autoregressive and moving average order of the INARMA model, the most general model among the possible models can be used for forecasting. This is especially useful for short history and high autocorrelation in data.

The findings of the thesis have been tested on two real data sets: (i) Royal Air Force (RAF) demand history of 16,000 SKUs and (ii) 3,000 series of intermittent demand from the automotive industry. The results show that for sparse data with long history, there is a substantial improvement in using INARMA over the benchmarks in terms of Mean Square Error (MSE) and Mean Absolute Scaled Error (MASE) for the one-step ahead forecasts. However, for series with short history the improvement is narrower. The improvement is greater for h-step ahead forecasts. The results also confirm the superiority of INARMA over the benchmark methods for lead time forecasts.

Acknowledgements

Completing a PhD is by far the most challenging yet rewarding task I have ever undertaken. Fulfilling this task was not possible without the support of numerous people, to whom I am grateful.

Most importantly, I would like to express my deep and sincere gratitude to my supervisor Prof. John Boylan, Head of Research, Faculty of Enterprise and Innovation, Buckinghamshire New University. His in-depth knowledge, constant encouragement, support and particularly his advice have been invaluable in ensuring I was on the right track throughout the PhD studies. I appreciate all his contributions of time and ideas.

I also wish to express my appreciation to Dr Aris Syntetos of Salford Business School, University of Salford, for providing me with the empirical data for this thesis. I would also like to thank Ms Laura Bray, Faculty Research Officer, Buckinghamshire New University for her sympathetic help she has provided during the last three years.

I am grateful to the Buckinghamshire New University for the scholarship they provided for this PhD study and also the support to enable me to attend and present my academic papers at international conferences.

My special appreciation goes to my parents, Mina Moradkhan and Gholamhossein Mohammadipour, who always encouraged me to concentrate on my studies. Their non-stop support has made my dreams come true. I also want to thank my brother, Dr Mani Mohammadipour, and my uncle, Dr Shahryar Mohammadipour, for their love and friendship.

Last but not least, words cannot express my special thanks to my beloved husband, Soroosh Saghiri, for his endless love, faith and encouragement in fulfilling this aspiration.

To Soroosh

my husband

my best friend

Contents

List of Illustrations

List of Tables

List of abbreviations

ACF	Autocorrelation Function
AE	Absolute Error
ARIMA	Autoregressive Integrated Moving Average
CLS	Conditional Least Squares
CML	Conditional Maximum Likelihood
DARMA	Discrete Autoregressive Moving Average
DSD	Discrete Self-Decomposable
EWMA	Exponentially Weighted Moving Average
GMM	General Method of Moments
IDF	Intermittent Demand Forecasting
INARMA	Integer Autoregressive Moving Average
LTD	Lead Time Demand
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MC	Modified Croston
MCMC	Markov Chain Monte Carlo
ME	Mean Error
MSE	Mean Square Error
ML	Maximum Likelihood
PACF	Partial Autocorrelation Function
PB	Percentage Better
RGRMSE	Relative Geometric Root Mean Square Error
SBA	Syntetos-Boylan Approximation
SBJ	Shale-Boylan-Johnston
SES	Single (or simple) Exponential Smoothing
SMA	Simple Moving Average
YW	Yule Walker
ZF	Zero Forecast

Chapter 1 INTRODUCTION

This chapter lays the foundations for this thesis and provides an overall perspective for this PhD research. It introduces the research by means of an overview of the project. The business context and research background are described. The research problem and research questions are then defined and the methodology is briefly outlined.

All the above topics are explained in detail by a step-by-step process through chapters 2 to 9. However, the introductory chapter is intended to outline the research through a summary of the research background and problems, expected results, and designated methodology. The structure of chapter 1 and sequences of sections are shown in Figure 1-1.

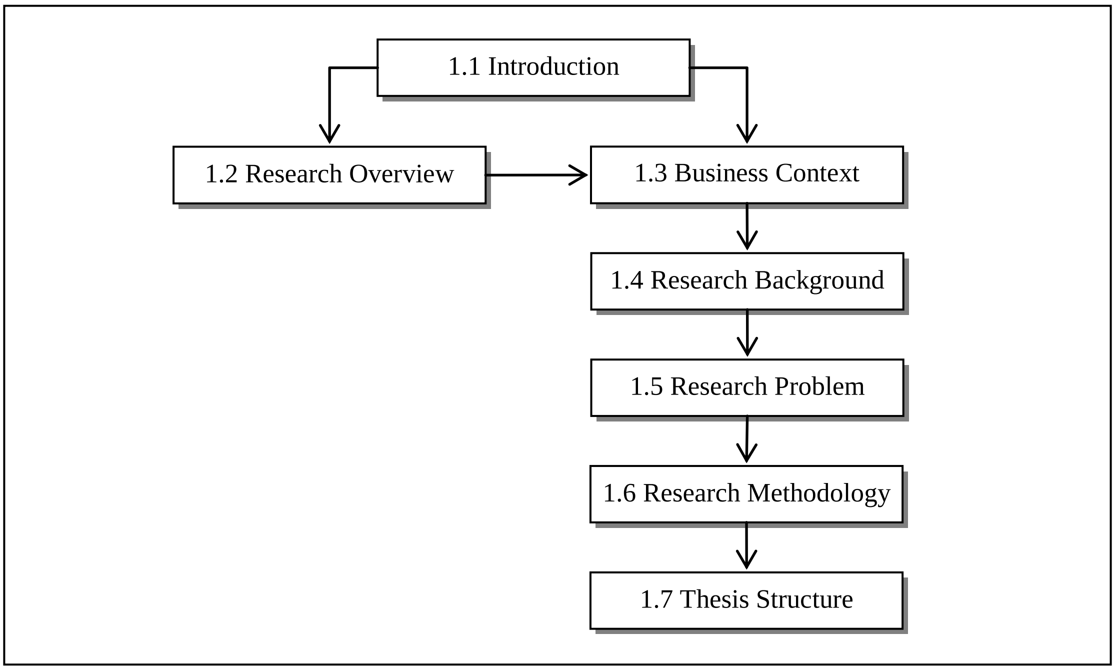

Figure 1-1 The structure of Chapter 1

1.1 Introduction

To attain a unified understanding of related concepts in this PhD thesis, a brief description of the key terms and phrases used in the study is provided in this section. These are the definitions that will be adopted in this thesis. More discussion will follow in later chapters.

Time series

Bowerman et al. (2005) have defined a time series as "*a chronological sequence of observations on a particular variable*" (p.4). Demand of a product over time and inventory level for a product over time are examples of time series.

One of the most general classes of models for forecasting a time series is the class of autoregressive integrated moving average (ARIMA) models (Box et al., 1994). ARIMA models are based on adding linear combinations of lags of the differenced series and/or lags of the forecast errors to the prediction equation, as needed to remove any last traces of autocorrelation from the forecast errors.

Time series of counts

Time series of counts is defined by McKenzie (2003) as "*counts of events, objects or* *individuals in consecutive intervals or at consecutive points in time*". Integer autoregressive moving average (INARMA) models have been proposed for forecasting time series of counts, and have received a certain amount of attention over the last 25 years. These models are explained in detail in chapter 3.

Intermittent series

Intermittent series are time series of non-negative integer values where some values are zero (Shenstone and Hyndman, 2005). A common example is intermittent demand. Intermittent demand should be distinguished from lumpy demand where the nonzero values are highly variable. Methods for intermittent demand forecasting (IDF) are reviewed in chapter 2.

1.2 Research Overview

This PhD concentrates on modelling intermittent demand by integer autoregressive moving average (INARMA) processes and proposing a forecasting method based on such processes. The study first focuses on stochastic characteristics of INARMA processes, including finding the second unconditional moment and the autocorrelation function (ACF) and partial autocorrelation function (PACF) structure. Lead time forecasts are also developed for INARMA models.

INARMA models are then used for intermittent demand forecasting. The results are compared to some benchmark methods in the literature, namely Croston's method (Croston, 1972), Syntetos-Boylan Approximation (SBA) (Syntetos and Boylan, 2005), and Shale-Boylan-Johnston (SBJ) (Shale et al., 2006). The rationale for this choice of benchmarks is given in chapter 2.

1.3 Business Context

Accurate demand forecasting is a significant concern for many organizations. It lays a foundation for every part of inventory management. Various forecasting models have been developed to incorporate different components of demand such as trend and seasonality. A class of demand called intermittent demand exists in which some periods have no demand at all. This is especially the case for service parts and capital goods (Willemain et al., 2004). Examples are commonly found in the aerospace, automotive, computer components, electronics, and industrial machinery industries. In general, all companies that must stock spare parts face intermittent demand.

Intermittent time series are also common in business and economic data. There are many cases for which non-intermittent data is intermittent at lower levels of data disaggregation (for example, greater frequency or smaller geography).

The items with intermittent demand are not rare and in fact they constitute the majority of items held by many stockists (Johnston et al., 2003). These items are also important from a financial point of view. In the U.S. alone, service parts management grew to a $700 billion business sector in 2001 (Patton and Steele, 2003). The fact that, in many cases, service parts face a high risk of obsolescence makes accurate forecasting for such items even more important. Lower stock-holding costs and higher service levels are the outcomes of more accurate forecasts.

INARMA models have had applications in a wide area including medical science (Franke and Seligmann, 1993; Cardinal et al., 1999), economics (Böckenholt, 1999; Brännäs and Hellström, 2001; Freeland and McCabe, 2004b) and service industries (Brännäs et al., 2002). However, the performance of these models in an intermittent demand context is yet to be tested. This testing is to be conducted in this research as a new application for INARMA models.

Although the empirical analysis of this thesis focuses on intermittent demand data, the theoretical and simulation findings can generally be applied to any time series with low non-negative integer-valued data.

1.4 Research Background

Many companies use single exponential smoothing (SES) for intermittent demand forecasting (Teunter and Sani, 2009) but Croston (1972) shows that this can result in biased forecasts and excessive stock levels. He proposes a method based on separate exponential smoothing estimates of demand size and interval between demands, which he claims to be unbiased. However, Syntetos and Boylan (2001) show that Croston's method is positively biased. Some methods have been suggested to reduce this bias (Levén and Segerstedt, 2004; Syntetos and Boylan, 2005). While it is confirmed that the correction by Syntetos and Boylan does reduce the bias of Croston's method, the modification by Levén and Segerstedt is even more biased than Croston's method (Teunter and Sani, 2009).

In parallel with these studies, some authors focus on finding a model underlying Croston's method (Shenstone and Hyndman, 2005; Snyder and Beaumont, 2007). This is especially useful to find the distribution and prediction intervals of the forecasts. However, none of these studies has yet demonstrated an underlying model for which Croston's method is optimal. It has been suggested that the focus should now be moved to stationary models for IDF such as time series models for counts (Shenstone and Hyndman, 2005).

Time series models for counts occur as counts of individuals (e.g. the number of people in a queue waiting to receive a service at a certain moment) or events (e.g. the number of accidents in a firm each three months). If these discrete variates are large numbers, they could be approximated by continuous variates; otherwise, special models should be used.

A class of models for count data has been developed, namely the integer autoregressive moving average (INARMA) models. These models were originally introduced in the 1980s (McKenzie, 1985; Al-Osh and Alzaid, 1987) and it will be shown in chapter 3 that they are analogous to well-known ARMA models (Box et al., 1994).

Intermittent demand data belong to a broader class called count data. Although count data frequently occur in many industries, not many stochastic models have been developed for them (Gooijer and Hyndman, 2006). As one of these few models, integer autoregressive moving average models have recently received much attention. This PhD thesis aims to bridge the gap between models for counts and intermittent demand forecasting, focusing specifically on INARMA models. In doing so, some issues such as identification, estimation of parameters and lead time forecasting need to be addressed. Identification of the autoregressive and moving average orders of INARMA models has been mainly conducted in the literature using the autocorrelation function (ACF) and the partial autocorrelation function (PACF) (Jung and Tremayne, 2006a; Zheng et al., 2006; Zhu and Joe, 2006; Bu and McCabe, 2008). To our knowledge, the ACF of the general INARMA(p,q) process has not been looked at before. Therefore, we establish the unconditional variance and the ACF of such processes. The conditional expected value of the over-lead-time aggregated process are also obtained. The accuracy of h-step-ahead and lead time INARMA forecasts are compared to those obtained by the benchmark methods mentioned in section 1.2. In doing so, the difficulties of forecasting intermittent demand are borne in mind, particularly if the length of data history is short or the data is sparse.

1.5 Research Problem

1.5.1 Initial Problem

The main problem addressed in this research is as follows:

- In the context of intermittent demand, is there any benefit in modelling and forecasting the demand using INARMA models, in terms of forecast accuracy, compared to simpler methods?

1.5.2 Research Questions

Based upon the above initial research problem, the four detailed questions for the research are as follows:

I. How can the appropriate integer autoregressive moving average (INARMA) model be identified for a time series of counts?

Different methods have been suggested for identification of ARIMA processes including: using the sample autocorrelation function (SACF) and the sample partial autocorrelation function (SPACF), and using a penalty function such as the Akaike information criterion (AIC) or the Bayesian information criterion

(BIC). This research focuses on finding the ACF and PACF structure of INARMA processes. But, using ACF and PACF needs a visual check. Identification based on these functions has not been automated for INARMA models as it has been for ARMA models (e.g. in AutoBox). Therefore automated methods should be used. It has been suggested in the literature that data should be first tested for any serial dependence. As a result, two identification procedures are compared in this research: a two-stage procedure first uses a test of serial dependence to distinguish between an independent and identically distributed (i.i.d.) process and other INARMA processes and then the AIC is used to select among other possible INARMA processes. A one-stage procedure does not have the first step and only uses the AIC. However, the correct model might not be identified at all times. In such cases, it is of interest to find the impact that misidentification has on forecast accuracy.

II. How can the parameters of integer autoregressive moving average (INARMA) models be estimated?

The performance of different estimation methods are to be tested in terms of both the accuracy of parameter estimates and their impact on forecast accuracy.

III. How can an INARMA process be forecasted over a lead time?

One of the application areas of lead-time aggregation is in the inventory control field where there is a lead time between placing an order by a manufacturer and receiving it from its supplier. The manufacturer has to place an order to cover the demand over the lead time and, therefore, the lead-time demand has to be forecasted. The aggregated INARMA(p,q) process over a lead time and its conditional expected value are found. The latter is then used as the lead time forecast.

IV. Do INARMA models provide more accurate forecasts for intermittent demand than non-optimal smoothing-based methods?

This research has suggested using INARMA models to forecast intermittent demand. The accuracy of forecasts provided by INARMA methods then has to be compared with some of the methods that have been used in the literature of intermittent demand forecasting. As this research solely focuses on demand forecasting and not inventory control, forecast-accuracy metrics will be used rather than accuracy-implication metrics (measures that analyze the effect of forecasting methods on inventory performance measures (Boylan and Syntetos, 2006)). As previously mentioned, the methods that will be used for comparison purposes are Croston's method (Croston, 1972), the Syntetos-Boylan Approximation (Syntetos and Boylan, 2005), and the method of Shale-Boylan-Johnston (Shale et al., 2006).

1.6 Research Methodology

This section will try to sufficiently clarify the “philosophy”, “approach” and “strategy” of this study to demonstrate how the expected results in this PhD may be achieved. This research follows the “positivism” philosophy where the research approach will be “deductive”. Accordingly, based upon the detailed hypotheses and the required level of generalisation, “simulation” and then “empirical study” is believed to be the most suitable strategy for this research.

To explain the research methodology of this study, the approach of the “research process onion” as termed by Saunders et al. (2003) is followed. On this basis, an appropriate research philosophy, research approach, and research strategy are proposed for the study. Consistent with the research strategy, source(s) of data are identified for analysis. The structure of this section is shown in Figure 1-2.

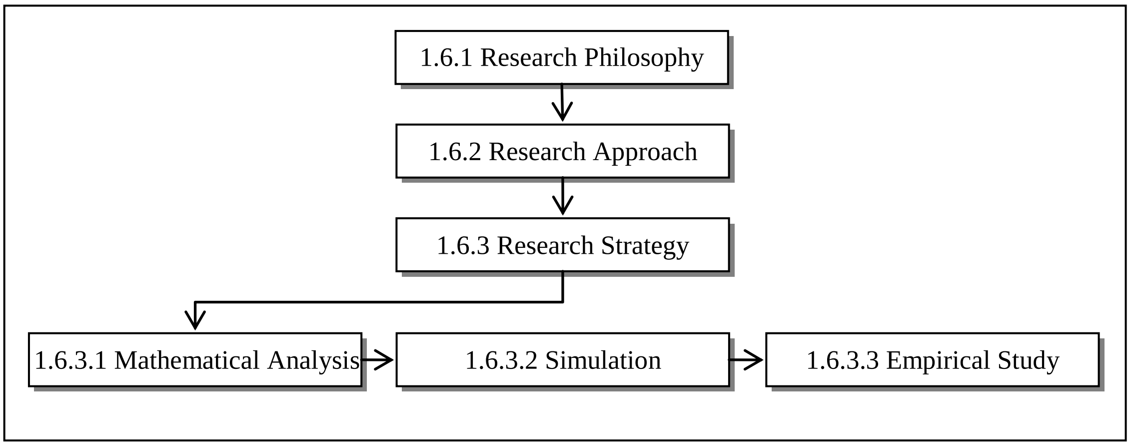

Figure 1-2 The structure of the research methodology

1.6.1 Research Philosophy

Taking into account the existing knowledge of the overall environment of forecasting and time series analysis, the ground philosophy of this research tends towards “positivism” (rather than pure “realism” or “interpretivism”), i.e. the end product of the research can be law-like generalisations. The nature of intermittent data makes it difficult for humans to detect patterns for this type of data. This means that a positivistic deductive approach is a natural starting point. An interpretivistic approach may be more appropriate for investigating human adjustments to intermittent demand forecasts.

1.6.2 Research Approach

Research approaches are classified into two main groups of “deductive” and “inductive”. While the former works on the basis of a clear understanding of the research theory and questions, the latter tries to find these through investigations with real world data (Saunders et al., 2003). Although some inductive studies have been done in the area of forecasting (e.g. Collopy and Armstrong, 1992; Adya et al., 2001), most of the studies in this field are deductive. A theory is developed first and a research strategy to test it is designed. This research is also based on developing a theoretical structure and testing the findings by simulation and empirical analysis. Hence, a deductive approach is followed in this PhD thesis.

1.6.3 Research Strategy

A research strategy endeavours to plan the process of answering the research questions (section 1.5.2). Our strategy consists of three steps: mathematical analysis, simulation and empirical study.

1.6.3.1 Mathematical Analysis

Although the introduction of INARMA models dates back to the 1980s, there are still an increasing number of studies which investigate the statistical properties of these models. The mathematical analysis of this research aims at extending the theory of INARMA modelling making it more complete.

The mathematical results of this PhD are provided in chapters 3 to 6. Some stochastic properties of the general INARMA(p,q) process have been obtained which are: the unconditional variance and the autocorrelation function (ACF) and partial autocorrelation function (PACF). The results of aggregation of an INARMA(p,q) process over lead time along with the conditional expected value of the aggregated process are established. These results are then used in chapters 8 and 9 to compare the performance of the INARMA method with some benchmark methods in forecasting intermittent demand.

1.6.3.2 Simulation

Simulation will be used for the following reasons:

- to assess the percentage of theoretically generated INARMA time series that can be identified correctly
- to investigate the effect of misidentification on forecasting accuracy
- to compare the performance of different estimation methods
- to assess the sensitivity of the results (parameter estimates and forecasting accuracy) to the sparsity of data and the length of history
- to compare the INARMA forecasts with those of the benchmark methods

It can be seen that simulation is essential for the first three objectives (and the part of the fourth objective relating to parameter estimates) because only for theoretically generated data, the order of the INARMA model and the parameters are known. For the remaining objectives, simulation is useful to gain additional insights. The design of the simulation study is explained in chapter 7.

1.6.3.3 Empirical Analysis

The findings of this PhD thesis are to be tested on real empirical data to assess the practical validity and applicability of the main results of the study. As discussed, simulation helps us to test the accuracy of model identification, when we know that the intermittent demand follows an INARMA process with known order and parameters. However, this is not true for real data in that we do not have such information from the beginning. Therefore, empirical analysis would help us to test the applicability of the results in real situations.

Although many studies focus on the statistical aspects of INARMA models, there are fewer studies regarding the application of these models (Jung and Tremayne, 2003). There is especially a lack of empirical testing of INARMA models on intermittent demand. The demand data series used in this PhD thesis are Royal Air Force (RAF) individual demand histories of 16,000 SKUs over a period of 6 years (monthly observations) and 3,000 series of intermittent demand for 24 periods (two years monthly series) from the automotive industry.

1.7 Thesis Structure

Based on the position adumbrated in this chapter, this PhD thesis is structured as follows. Chapter 2 discusses different definitions and categorizations of intermittent demand, reviews methods for intermittent demand forecasting and the accuracy measures that can be used in this context.

Chapter 3 briefly reviews different count models and introduces INARMA models. The stochastic properties of these models and their application in the literature are examined. The unconditional variance and ACF of the INARMA(p,q) process are obtained.
Chapter 4 reviews different approaches for identification of the order of INARMA models. The PACF structure of INARMA models is also derived.
Chapter 5 discusses methods for estimation of the parameters of these models.
Chapter 6 investigates the forecasting of an INARMA process over lead time.
Chapter 7 discusses the design of simulation, and the results of simulation experiments

are presented in Chapter 8.

Chapter 9 assesses the validity of theoretical and simulation results on real intermittent demand data.

Finally, the findings of this PhD research, the limitations and some potential future studies are reviewed in chapter 10. The structure of the thesis is shown in Figure 1-3.

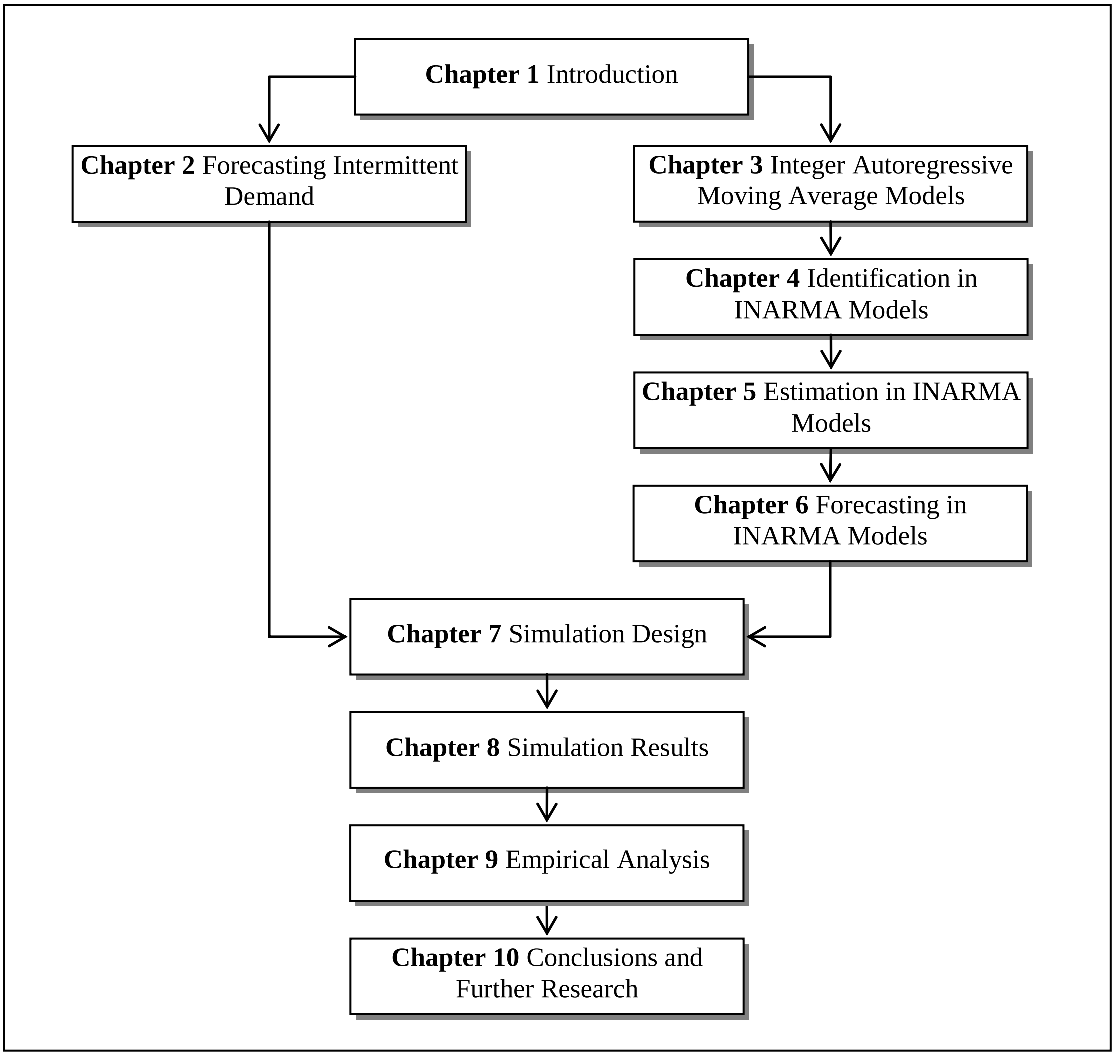

Figure 1-3 The structure of the thesis

Chapter 2 FORECASTING INTERMITTENT DEMAND

2.1 Introduction

This chapter aims to provide an overview of the literature on forecasting intermittent demand. This research focuses on integer autoregressive moving average (INARMA) models to address intermittent demand modelling and forecasting. As the focus of this research is on forecasting, we do not review the literature on inventory control for slow-moving items.

The chapter is organized as follows. Intermittent demand is defined in detail in section 2.2. Methods for forecasting intermittent demand are then discussed in section 2.3. We start our review with Croston's method (Croston, 1972) as the most widely used approach in this field. Some variants of Croston's method are then reviewed. Studies based on bootstrapping to forecast lead-time demand are also discussed. As this research is based on comparing different forecasting methods, the forecast accuracy measures need to be selected. Some accuracy measures cannot be used in an intermittent demand context. Section 2.4 determines these measures and classifies the measures that can be applied to intermittent series. A number of studies have compared different methods of forecasting intermittent demand. These studies are reviewed in section 2.5. The motives to use INARMA models for modelling and forecasting intermittent demand are discussed in section 2.6 and, finally, section 2.7 concludes the chapter.

2.2 Definition of Intermittent Demand

Intermittent demand is defined by Silver et al. (1998) as "*infrequent in the sense that the average time between transactions is considerably larger than the unit period, the latter being the interval of forecast updating*" (footnote, p.127). The main disadvantage of this definition is its impracticality, i.e. it does not define how long the average time between transactions should be for demand to be considered intermittent.

The definition provided by Johnston and Boylan (1996) is more practical. They suggest that demand is intermittent when the mean inter-arrival time between demands is greater than 1.25 review intervals. This cut-off value is based on a comparison of Croston's method and Single Exponential Smoothing (SES) using simulated data. However, this definition also has its limitations: it is simulation based, it depends on specific methods, and it does not take into account the "lumpiness" of the demand. Intermittent demand is often lumpy, which means that the variability among the nonzero values is high (Willemain et al., 2004). However, these two concepts should be distinguished.

Shenstone and Hyndman (2005) introduce another practical definition: "Data for intermittent demand items consist of time series of non-negative integer values where some values are zero". It can be seen that this does not take into account the lumpiness of demand either.

There are two main categorization schemes for intermittent demand in the literature. The first approach, proposed by Williams (1984) and later revised by Eaves and Kingsman (2004), is based on second moment variability (transaction variability, demand size variability and lead-time variability). The other approach, suggested by Syntetos et al. (2005) is based on frequency of transactions and demand size variability.

The classification approach by Eaves and Kingsman (2004) is a revision of the method of Williams (1984) who classified demand by decomposing the variance of lead-time demand into transaction variability, demand size variability and lead-time variability. This is shown in Table 2-1.

Table 2-1 The categorization scheme for intermittent demand data (Eaves and Kingsman, 2004)

Lead-time demand component			**Demand pattern classification**
Transaction variability	*Demand size variability*	*Lead-time variability*	
Low	Low		Smooth
Low	High		Irregular
High	Low		Slow moving
High	High	Low	Mildly intermittent
High	High	High	Highly intermittent

The main disadvantage of Eaves and Kingsman's classification is its lack of practicality. Unlike Syntetos et al. (2005), they do not provide cut-off values for different classes.

Syntetos et al. (2005) categorize demand based on the expected mean square error (MSE) of each forecasting method. They propose four categories of demand shown in Figure 2-1 which are: "erratic", "lumpy", "smooth", and "intermittent". Each of these demand classes are uniquely specified by two parameters: p which is the average inter-demand interval, and υ which is the squared coefficient of variation of the demand when it occurs.

The classification by Johnston and Boylan (1996) is based on comparing Croston's method with SES as methods widely used in forecasting software packages. Therefore, although it has the benefit of being practical, one can argue that it is not a comprehensive classification method that can be used for all forecasting methods. The categorization scheme of Syntetos et al. (2005) also includes the Syntetos-

Boylan Approximation (SBA) (explained in section 2.3.1.1) in addition to the other two methods. The main benefit of their scheme is again its practicality (because of determining the cut-off values) and that it is empirically validated. However, as the authors mentioned, the categories are based on only one forecast accuracy measure (MSE) (see section 2.4 for more information on accuracy measures).

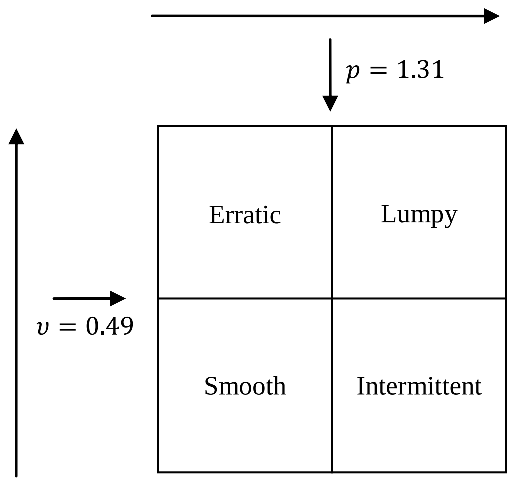

Figure 2-1 The categorization scheme for intermittent demand data (Syntetos et al. 2005)

The classification by Syntetos et al. (2005) distinguishes between intermittent and smooth demand to determine which method should be used in each situation. This is because the Syntetos-Boylan Approximation and SES methods are not universal, i.e. these methods are not appropriate for both categories of demand.

This research aims at using integer autoregressive moving average (INARMA) models to model and forecast intermittent demand. This method can be used for both intermittent and non-intermittent data provided that the data is not lumpy. Therefore, we do not need to focus on a specific classification method for intermittence to find when the method should be used. We do need to follow a standard definition that is practical and universal (not limited to specific methods).

Among the definitions that we reviewed before, only the one by Shenstone and Hyndman (2005) met all the above criteria and therefore is used for the purpose of this study. As a result, whenever a data series has at least one zero, it is considered as an intermittent data series, and an INARMA forecasting method can be used. However, the universality of the INARMA approach means that it can also be used when all the observations are positive.

As previously mentioned, although the classification by Syntetos et al. (2005) is practical, it is only based on specific forecasting methods. We will test this classification when the data is INARMA. INARMA forecasts will be compared to the best benchmark method based on this classification.

The main reason for not using Eaves and Kingsman classification method is its lack of practicality and also because lead-time variability will not be studied in this thesis. As mentioned before, the other reason for the selection of Syntetos et al.'s classification is that two methods used in their classification (Croston's method and SBA) will be used in this research for comparison reasons.

This research focuses on using INARMA models with Poisson marginal distributions to model intermittent demand. Due to the properties of the Poisson distribution, lumpy demand cannot be modelled by these models. As will be explained in chapter 9, a filtering mechanism will be used to eliminate the lumpy series when dealing with empirical data.

2.3 Methods of Forecasting Intermittent Demand

Inventories with intermittent demands are common in practice (Shenstone and Hyndman, 2005) and they create significant problems in the manufacturing and supply environment (Syntetos and Boylan, 2001). The accurate forecasting of demand is one of the most important issues of inventory management (Ghobbar and Friend, 2003). This is more difficult when demand has an intermittent nature (Willemain et al., 2004).

Willemain et al. (2004) divide intermittent demand forecasting (IDF) methods into:

- simple statistical smoothing methods
- Croston's method and its variants
- bootstrap methods

from which we focus on the last two classes because, according to Croston (1972), the first class results in biased forecasts when applied immediately after a demand occurrence. These methods have also been the subject of some comparison studies, including Teunter and Duncan (2009). These studies are reviewed in section 2.5.

2.3.1 Croston's Method

The most widely used approach for forecasting intermittent demand is Croston's method (Shenstone and Hyndman, 2005). Croston (1972) suggested that conventional forecasting methods like Simple Moving Averages (SMA) and Single Exponential Smoothing (SES) may not be appropriate for slow-moving items. The bias associated with SES is expressed in a quantitative form for the case in which forecasts are updated after a demand occurrence. He then proposed a method based on separate Single Exponential Smoothing (SES) forecasts on the size of a demand and on the time period between observing two demands, both with the same smoothing constant α.

Let Y_t be the demand occurring during the time period t and the indicator variable X_t be:

$$X_t = \begin{cases} 1 & \text{when demand occurs} \\ 0 & \text{when no demand occurs} \end{cases}$$

Equation 2-1

Furthermore, let j_t be the number of periods with nonzero demand during the interval $[0, t]$, $j_t = \sum_{i=1}^{t} X_i$. The size of the jth non-zero demand is then shown by Y_j and the inter-arrival time between Y_{j-1} and Y_j is Q_j.

Croston (1972) adopts a stochastic model of arrival and size of demand, assuming that demand sizes, Y_j, are normally distributed, $N(\mu, \sigma^2)$, and that demand is random and has a Bernoulli probability $1/p$ of occurring in every review period (subsequently, the inter-arrival time, Q_j, follows the geometric distribution with a mean p). Both demand sizes and intervals are assumed to be stationary.

The demand at period t is given by:

$$Y_t = X_t Y_j$$

Equation 2-2

Using Croston's method, the demand size and inter-arrival time between demands are separately forecasted using Single Exponential Smoothing (SES), with forecasts being updated only after demand occurrence. Let Z_j and P_j be the forecasts of the $(j+1)$th demand size and inter-arrival time respectively, based on demands up to period j. Then, based on Croston's method:

$$Z_j = (1-\alpha)Z_{j-1} + \alpha Y_j$$

Equation 2-3

$$P_j = (1-\alpha)P_{j-1} + \alpha Q_j$$

Equation 2-4

The forecast for the next time period is then given by the smoothed size of demand divided by the smoothed inter-arrival time:

$$\hat{Y}_j = Z_j / P_j$$

Equation 2-5

According to Croston (1972), the expected value and variance of the forecast for the stochastic model are given by:

$$E(\hat{Y}_j) = \mu/p$$

Equation 2-6

$$\mathrm{var}(\hat{Y}_j) = \frac{\alpha}{2-\alpha}\left(\frac{p-1}{p^2}\mu^2 + \frac{\sigma^2}{p}\right)$$

Equation 2-7

based upon which, he claims the method is unbiased. However, Syntetos and Boylan (2001) showed that Croston's method is biased due to the fact that $E(\hat{Y}_j) \neq E(Z_j)/E(P_j)$. Syntetos and Boylan (2005) suggest some modifications to reduce the bias which will be discussed in section 2.3.1.1.

Rao (1973) made corrections to some of the expressions of the forecast variance. However, these changes have no effect on the forecast of mean demand.

Croston's method is based on using the same smoothing constant (α) for updating demand sizes and demand intervals. Schultz (1987) suggests using different smoothing constants, α and β, for size and interval between demands.

Croston's method is also based on assumptions of independence and normality of demand, independence of demand sizes and inter-arrival times, and independence of inter-arrival times with a Geometric distribution. The validity of these assumptions for real-world data has been discussed by several authors (e.g. Willemain et al., 1994; Snyder, 2002; Shenstone and Hyndman, 2005). Willemain et al. (1994) argue that not only might there be autocorrelation among demand sizes and inter-arrival times, but it is also possible for sizes and intervals to be correlated.

Shenstone and Hyndman (2005) explore possible models underlying Croston's method. These models are reviewed in section 2.3.1.4.

2.3.1.1 Bias Correction for Bernoulli Demand Incidence

As explained in the previous section, Syntetos and Boylan (2001) showed that Croston's forecasts are biased. A new estimator is proposed later by Syntetos and Boylan (2005) to reduce the bias associated with Croston's method. It is shown that the bias can be approximated by $\left(\frac{\alpha}{2-\alpha}\mu\frac{p-1}{p^2}\right)$ as:

$$E\left(\frac{Z_j}{P_j}\right) \approx \frac{\mu}{p} + \frac{\alpha}{2-\alpha}\mu\frac{p-1}{p^2}$$

Equation 2-8

and based on Equation 2-8 they suggest that the forecast should be multiplied by $(1-\frac{\alpha}{2})$ in order to reduce the bias.

$$E\left(\left(1-\frac{\alpha}{2}\right)\frac{Z_j}{P_j}\right) \approx \frac{\mu}{p}$$

Equation 2-9

Therefore, the new estimator of mean demand is:

$$\hat{Y}_j = \left(1-\frac{\alpha}{2}\right)\frac{Z_j}{P_j}$$

Equation 2-10

All of Croston's assumptions are maintained in the derivation of the new estimator (however, the assumption of normality of demand sizes is not necessary for the derivation). Some of these assumptions have been changed in another study by Shale et al. (2006) which will be discussed in the next section.

The new method, called the Syntetos-Boylan Approximation (SBA) in the literature, is then compared to three other forecasting methods: Simple Moving Average, Single Exponential Smoothing, and Croston’s method on 3,000 real intermittent demand data series from the automotive industry. The results suggest that the new method is the most accurate estimator for the faster intermittent demand data.

Teunter and Sani (2009) also compared Croston's method, SBA and modified Croston (see section 2.3.1.3) and show that SBA has the smallest average standard deviations of all.

2.3.1.2 Bias Correction for Poisson Demand Incidence

Shale et al. (2006) consider the case of Poisson demand incidence instead of a Bernoulli distribution as in Croston’s model. Also, they assume that the inter-arrival time follows a negative exponential instead of a geometric distribution. They derive the bias expected for this case and provide the correction factor for application. This is done for the cases where either a simple or an exponentially weighted moving average (EWMA) is used. Hereafter, this method is called the Shale-Boylan-Johnston (SBJ) method.

For the case of a simple moving average, when the k most recent inter-arrival times have a negative exponential distribution, the average of them has an Erlang distribution. The correction factor for an adapted Croston’s estimate of mean demand for this case is shown to be $[(k-1)/k]$. The adaptation consists of a ratio of simple moving average of demand size and demand interval. The estimate of demand would then be:

$$E(\hat{Y}_j) = \left(\frac{k-1}{k}\right)\frac{E(Z_j)}{E(P_j)}$$

Equation 2-11

where $E(Z_j)$ and $E(P_j)$ are the arithmetic averages rather than the exponentially weighted averages of demand sizes and interval between demands. In the latter case, it is shown that the smoothing parameter (A) of EWMA is linked to k as follows:

$$k = \frac{2 - A}{A}$$

Equation 2-12

Based on the above result, the estimate of demand (which will be used later in this research) is:

$$E(\hat{Y}_j) = \left(1 - \frac{A}{2 - A}\right)\frac{E(Z_j)}{E(P_j)}$$

Equation 2-13

When k is not an integer, the probability distribution is the continuous form of the Erlang distribution which is Gamma distribution. The correction factor for this case is also $[(k-1)/k]$.

It has been shown through simulation that these correction factors are very close to the observed average bias associated with Croston's method (Shale et al., 2006).

2.3.1.3 Modified Croston (MC)

Levén and Segerstedt (2004) propose a Modified Croston procedure, based on an earlier working paper by Segerstedt (2000), which they claim works for both fast-moving and slow-moving items. The procedure is given by:

$$\hat{d}_n^{MC} = \hat{d}_{n-1}^{MC} + \alpha\left(\frac{X_n}{T_n - T_{n-1}} - \hat{d}_{n-1}^{MC}\right)$$

Equation 2-14

where $\hat{d}_n^{MC}$ is the forecasted demand rate at the end of period T_n, T_n is the time period in which the demand X_n occurs, X_n is the measured demand quantity during the *n*th period, and α is the smoothing constant. Demand is assumed to follow the Erlang distribution. This modified Croston procedure is then compared to Croston's method in which:

$$\hat{d}_n^C = \frac{\hat{X}_n}{\hat{I}_n}$$

Equation 2-15

$$\hat{X}_n = \hat{X}_{n-1} + \alpha(X_{n-1} - \hat{X}_{n-1})$$

Equation 2-16

$$\hat{I}_n = \hat{I}_{n-1} + \alpha(I_{n-1} - \hat{I}_{n-1})$$

Equation 2-17

Where $\hat{d}_n^C$ is the forecasted demand rate using Croston's method and $\hat{I}_n$ is the forecasted demand interval calculated at the end of period T_n with other notations as before.

The simulation results suggest that an inventory control system based on the MC procedure and the Erlang distribution yields fewer shortages than a system using exponential smoothing and the Normal distribution. The authors believe that is due to the MC procedure and not on the assumption of Erlang distribution.

Levén and Segerstedt (2004) claim that the MC estimator avoids bias. However, Boylan and Syntetos (2007) prove that not only is it biased, but also its bias is substantially greater than the original Croston procedure, especially for highly intermittent series. The results of a comparison study by Teunter and Sani (2009) confirm the bias of the MC method, showing that it has the highest bias when compared to the original Croston's method, SB approximation and the Syntetos approximation (Syntetos, 2001). Therefore, the MC method will not be considered further in this thesis.

2.3.1.4 Models underlying Croston's Method

Shenstone and Hyndman (2005) use autoregressive integrated moving average (ARIMA) models for both size of demand and inter-arrival times. In an attempt to find an underlying model for Croston's method, they have looked at several models. First, they assume that $Y_j \sim \text{ARIMA}(0,1,1)$ and $Q_j \sim \text{ARIMA}(0,1,1)$, where Y_j is the size of the *j*th non-zero demand, Q_j is the inter-arrival time between Y_{j-1} and Y_j, and Y_j and Q_j are independent. They also restrict the sample space of the underlying model to be positive by defining $\log(Y_j) \sim \text{ARIMA}(0,1,1)$ and $\log(Q_j) \sim \text{ARIMA}(0,1,1)$. The other two models are the ones proposed by Snyder (2002) in which $Y_j \sim \text{ARIMA}(0,1,1)$ and $Q_j \overset{\text{i.i.d.}}{\sim} \text{Geometric}(p)$, and $\log(Y_j) \sim \text{ARIMA}(0,1,1)$ and $Q_j \overset{\text{i.i.d.}}{\sim} \text{Geometric}(p)$ where p is the average inter-arrival time of the series.

Shenstone and Hyndman (2005) suggest that any model assumed to be underlying Croston's method must be non-stationary and defined on a continuous sample space including negative values. However, they argue that none of the above mentioned models is consistent with the properties of intermittent demand data. Finally, they suggest that instead of focusing on models based on SES, it is worth considering stationary models such as Poisson autoregressive models. These models have not been used for intermittent demand data before, opening up a new area of study.

In a recent working paper, Snyder and Beaumont (2007) claim to find a logically sound model underlying Croston's method. They assume that the sizes of positive demands follow a Poisson distribution. However, they assume that the Poisson distribution is shifted by one to the right to avoid zero demands. This results in the following probability mass function for a positive demand at period t:

$$P_{tj} = \frac{(Z_{t-1} - 1)^{j-1}}{(j-1)!} e^{-Z_{t-1}+1}$$

Equation 2-18

where j is the value of the positive demand. The probability of a positive demand in period t is then defined by π_t which is:

$$\pi_t = \frac{1}{P_t}$$

Equation 2-19

Finally, the probability of demand in period t is given by:

$$P(Y_t = j) = \begin{cases} 1 - \pi_t & \text{if } j = 0 \\ \pi_t P_{tj} & \text{if } j > 0 \end{cases}$$

Equation 2-20

The above equation is claimed to be *the* underlying model for Croston's method

(Snyder and Beaumont, 2007). It can be considered as *a* model underlying Croston's method. However, they have not performed any assessment of the optimality of the proposed model for Croston's method. If a model exists for which Croston’s method is optimal, then it has yet to be formally established and proven.

2.3.2 Bootstrapping

Bootstrapping has been proposed by Willemain et al. (2004) and Porras (2005). It is based on repeatedly sampling L demands from demand history to estimate the distribution of lead time demand (lead time = L). The main advantage of bootstrapping is that it forecasts the whole lead time demand distribution.

There are many variants of the bootstrapping method (see for example Efron, 1979; Bookbinder and Lordahl, 1989) many of them having the disadvantage of being complex, a disadvantage that also holds for the bootstrapping method proposed by Willemain et al. (2004). The bootstrapping method proposed by Porras and Dekker (2007) is simpler. Both of these methods along with the method proposed by Snyder (2002) are discussed in this section.

2.3.2.1 Snyder (2002)

Snyder (2002) proposes a parametric bootstrap method to generate an approximation for the lead-time demand distribution from Croston's model. In each iteration, first, the values for noise terms are generated from a normal distribution with mean 0 and variance σ^2. Then the values for indicator variables (X_t, Equation 2-1) are generated from a Bernoulli distribution with probability p. A realisation of future demand series is then produced based on Croston’s method (Equation 2-3 and Equation 2-4), and finally, the lead-time demand is calculated from $\sum_{t=n+1}^{n+L} Y_t$.

In order to tackle the problem of having negative demands, Snyder (2002) suggests two adaptations. The first is to apply exponential smoothing to the logarithm of the data, which is called *log-space adaptation*. The other method, called the *adaptive variance version*, differs from Croston’s method in two respects: (*i*) variability is measured in terms of variances instead of mean absolute deviations (MAD); (*ii*) a second smoothing parameter, β, is used to define changes of variability over time.

Snyder (2002) claims that the advantage of the methods is that they can be applied to both fast and slow moving demand. However, they have the disadvantages of underestimating the variability of lead-time demand due to ignoring the effects of estimation error, and being based on the Normal distribution, which may not be appropriate when demand sizes are small.

2.3.2.2 Willemain et al. (2004)

Bootstrapping (Efron, 1979) produces pseudo-data by sampling with replacement from the observations. Willemain et al. (2004) develop a modified bootstrap to forecast the distribution of the sum of intermittent demands over a fixed lead time. The modification allows for autocorrelation, frequent repeated values, and relatively short series which are ignored in conventional bootstrapping.

The underlying model of intermittent demand incidence that they assume is a two-state, first order Markov process.

As for the demand sizes, two models have been discussed. The first *model* only assumes that demand sizes are stationary and they can be obtained by the *method* of sampling from the nonzero values that have appeared in the past. The problem here, as also mentioned by Willemain et al. (2004), is that no different values would appear in the future. As a result, they suggest an ad hoc *method* to deal with demand sizes called “jittering” based on no *model* for demand sizes.

The method is based on generating a sequence of zero and nonzero values over the L periods of the lead time. The state transition probabilities are estimated from the historical demand series using started counts. Then values are assigned to nonzero forecasts. Here, instead of only choosing nonzero values that have appeared in the past, which results in having no different values in the future, they *jitter* the selected value from the past, i.e. add some random variation to it.

Although some methods for intermittent demand forecasting are based on the assumptions that demands in each time period are independent and normally distributed, with neither of them necessarily being valid for intermittent demand (Willemain et al., 2004), Willemain et al.'s bootstrap method requires neither assumption. It only assumes that demand is stationary.

The main advantage of bootstrapping is that it does not rely on any distribution while all other methods do. However, the model of Willemain et al. (2004) for demand occurrence is not general and also the method of producing variable demand sizes is ad hoc.

They compare the accuracy of the developed bootstrap against exponential smoothing and Croston's method by applying them to over 28,000 items provided by nine industrial companies. The results suggest that, although Croston's method provides more accurate estimates of the mean level of demand at the moments when demand occur, it does not outperform exponential smoothing when forecasting the lead time distribution. However, the bootstrap provides an improvement on exponential smoothing, especially for short lead times.

Gardner and Koehler (2005) argue that Willemain et al. (2004) did not use the correct lead time demand distribution for exponential smoothing and Croston's method in their comparison. They also argue that Willemain et al. (2004) should have considered the modifications to the Croston's method that have shown improvements.

2.3.2.3 Porras and Dekker (2007)

Porras and Dekker (2007) propose another procedure to specify the lead time demand (LTD) distribution. Known as the Empirical Method, it differs from Willemain's in that it constructs a histogram of demand over lead time without sampling. Therefore, it has the benefit of capturing the autocorrelation of LTD. It is also easier than Willemain's bootstrap to implement. However, when lead time is long and the length of demand history is short, which is often the case for intermittent demand data, there are few blocks of LTD to select from.

Another problem with this method is that, particularly for short lead times (say $L =$

1), it is not possible to attain high percentiles in the demand distribution. Even if a high percentile can be reached, it might be an outlier and not representative of the population. Willemain et al. (2004) tackle this issue by introducing jittering, as explained in the previous section.

An empirical study is conducted to compare the performance of the new method from an inventory control perspective to other methods including: Willemain's bootstrap, Normal distribution, and Poisson distribution. For the normal-based model, it is assumed that demand per period follows a Normal distribution. The average and standard deviation of the observed demand are then used to estimate the parameters of the Normal lead time demand. When the lead time demand is assumed to have a Poisson distribution, the parameter is estimated from the demand data for the different items.

The results show that although demand of spare parts does not generally follow a Normal distribution, the Normal model performs well, producing the highest savings in inventory costs. It has also been found that the Empirical Method outperforms the Willemain method.

2.3.3 Causal Models

As pointed out by Hua et al. (2007), demand of spare parts can be attributed both to the status of the equipment and the spare parts, and to the maintenance policy. Hua et al. (2007) establish the impact of the maintenance policy through finding the relationship between explanatory variables and the nonzero demand of spare parts. They develop a method called the Integrated Forecasting Method (IFM) that first forecasts the occurrence of nonzero demand and then estimates the lead time demand. The first is done using an autocorrelation function to choose either a Markov process or explanatory variables, while the second is done by sampling from the nonzero values observed in the past and summing them over the lead time. They compare the results of the Markov Bootstrapping (MB) method (same as Willemain's bootstrapping method except that the jittering technique is not used) with IFM. The performance of SES, Croston's method, MB and IFM is also compared based on the mean absolute percentage error of lead time demand

(MAPELTD). The results confirm that Croston's method provides more accurate estimates of mean LTD than the other methods.

Ghobbar and Friend (2002) study the source of lumpiness of demand for aircraft maintenance repair parts, in order to reduce the occurrence of part shortages. They study the effect of five environmental factors on lumpiness. These factors are: the primary maintenance process (PMP) (including hard-time and conditional-monitoring), the aircraft utilization rate (AUR), the component's overhaul life (COL), the squared coefficient of variation of demand (CV^2) and the average inter-demand interval (ADI). The first three factors are independent variables and the last two are dependent variables in the general linear model (GLM). The results show that the demand variability increases when the level of aircraft utilization and flying hours increases. It shows that AUR can be a major source of lumpiness because it increases the CV^2 and decreases the ADI for the observed demand.

In another study, they compare the results of 13 forecasting methods including Croston's method, SES, and also those used by aviation companies (Ghobbar and Friend, 2003). Four environmental factors considered in this study are: the seasonal period length, primary maintenance process (PMP), squared coefficient of variation (CV^2) and the average inter-demand interval (ADI). CV^2 and ADI are taken as covariate factors while the other two are considered as categorical factors. The variation attributable to each factor and their interactions is studied through using analysis of variance (ANOVA).

Ghobbar and Friend (2003) also establish a predictive error-forecasting model (PEFM) to compare different forecasting methods based on their factor levels to evaluate which method is the best in any situation. The model is based on a GLM that predicts a response variable using its relationship with factor variables.

The results of studies on causal models show an interesting line to pursue. These models are especially useful when a short demand history is available and time series methods cannot be used (Boylan and Syntetos, 2008). However, these models have not yet been well developed in the literature. As a future line of study, the effect of incorporating these models and INARMA models on the accuracy of forecasts can be investigated (see chapter 10).

2.3.4 Conclusions on IDF Methods

Three classes of methods for intermittent demand forecasting have been reviewed in this section, namely Croston's method and its variants, bootstrapping methods, and causal methods.

Empirical studies have proved the ability of Croston's method and variants (especially SBA) to produce reasonable forecasts (Eaves and Kingsman, 2004; Syntetos and Boylan, 2005). However, properties of these methods (such as bias) have been derived based on assumptions of independence of demand, independence of demand sizes and inter-arrival times. As mentioned, some of these assumptions are not valid for real-world data. Another problem with Croston's method is that it has not yet been shown to be optimal for a specific demand model. Modelling intermittent demand by INARMA models makes it possible to take into account the correlation between demands. Another advantage is that optimal methods are known to exist for these models.

Bootstrap methods have the advantage of having no distributional and independence assumptions. The main disadvantage of these methods is that they are rather complex. These methods along with causal methods have not been well developed in the literature and more comparative studies with the best benchmark methods (as criticized by Gardner and Koehler (2005)) need to be done to prove their performance.

2.4 Assessing Forecast Accuracy

The nature of intermittent demand data makes some of the conventional accuracy measures inappropriate. For example, when one or more of the observed demands is zero, the mean absolute percentage error (MAPE) is undefined and, therefore, cannot be used.

In fact, none of the relative-to-the-series accuracy measures can be used because zero observations may yield "division by zero" problems. This includes MAPE, Median Absolute Percentage Error (MdAPE), Root Mean Square Percentage Error (RMSPE), and Root Mean Square Percentage Error (RMSPE).

Although the symmetric MAPE introduced by Makridakis and Hibon (1995) tackles this issue by dividing the absolute error by the average of the actual observation and the forecast:

$$\text{sMAPE} = \sum_{t=1}^{N} \frac{Y_t - F_t}{(Y_t + F_t)/2} / N \times 100$$

Equation 2-21

it is known that it suffers from asymmetry problems in its treatment of negative and positive errors (Goodwin and Lawton, 1999). Also, if the actual value Y_t is zero, sMAPE would always be 200%, which is meaningless (Syntetos, 2001).

As discussed by Syntetos and Boylan (2005), all relative-to-a-base accuracy measures that relate the forecast error to a benchmark, usually naïve 1 method, should also be discarded because the error would often be zero.

The mean absolute scaled error (MASE) proposed by Hyndman and Koehler (2006) does not have the problems seen with previous measures. This measure is obtained by scaling the absolute error based on the *in-sample* MAE from a benchmark forecast method. Assuming the benchmark method is the naïve method, the MASE is defined as:

$$\text{MASE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{Y_t - F_t}{\frac{1}{N-1} \sum_{i=2}^{N} |Y_i - Y_{i-1}|} \right|$$

Equation 2-22

A $\text{MASE} < 1$ means that the forecasting method is on average better than the naïve forecasts, and a $\text{MASE} > 1$ indicates that the method is on average worse than the naïve method. As explained by Hyndman (2006), the in-sample MAE is always available and more reliably non-zero than any out-of-sample measures. The MASE would be infinite or undefined only if all historical observations are equal. However, using in-sample MAE has a disadvantage of making MASE vulnerable to outliers in the historical time series (Kolassa and Schütz, 2007). As a result, the MASE of two time series with the same forecasts and identical true demands during the forecast horizon will be different if the two series differed in their historical demands. This makes it a more complicated metric to interpret.

Syntetos and Boylan (2005) categorize the accuracy measures for the purpose of comparing methods in an intermittent demand context into two categories: absolute accuracy measures and accuracy measures relative to another method. Each of these categories is reviewed as follows.

2.4.1 Absolute Accuracy Measures

These measures are calculated as a function of the forecast errors alone. Examples include the mean square error (MSE) and the mean absolute error (MAE). Theoretically, these measures can be computed for intermittent demand. However, when averaged across many time series, they do not take into account the scale differences between them (Syntetos and Boylan, 2005). Mean error (ME) can be considered as an exception to the above rule because it takes into account the sign of the error and is less susceptible to scale effects. ME is given by:

$$\mathrm{ME} = \sum_{t=1}^{N} \frac{Y_t - F_t}{N}$$

Equation 2-23

If ME is divided by the average demand per unit time period, the scale dependencies are eliminated.

2.4.2 Accuracy Measures Relative to another Method

These measures are calculated as a ratio to other forecasting methods. Examples include percent better (PB) and the relative geometric root mean square error (RGRMSE).

The percentage better measure counts and reports the percentage of time that a given method has a smaller forecasting error than another method.

The RGRMSE for methods A and B in a time series is:

$$\text{RGRMSE} = \frac{\left(\prod_{t=1}^{N}(Y_t - F_{A,t})^2\right)^{\frac{1}{2N}}}{\left(\prod_{t=1}^{N}(Y_t - F_{B,t})^2\right)^{\frac{1}{2N}}}$$

Equation 2-24

The theoretical properties of RGRMSE are discussed by Fildes (1992). He assumes that the squared errors of a particular series have the form of:

$$e_{M,t+L}^2 = \varepsilon_{M,t+L}^2 . \upsilon_{t+L}$$

Equation 2-25

where $(e_{M,t+L} = Y_{t+L} - F_{M,t+L})$, υ_{t+L} are assumed positive and can be thought of as errors due to the particular time period affecting all methods equally, while $\varepsilon_{M,t+L}$ are the method's (M) specific errors. He argues that the model of Equation 2-25 represents the case where data (and therefore errors) are contaminated by occasional outliers. He then shows that the geometric (rather than arithmetic) RMSE is independent of υ_{t+L}.

It can be seen from Equation 2-24 that GRMSE is identical to GMAE because the square and the square root cancel each other (Hyndman, 2006):

$$\text{GRMSE} = \frac{\left(\prod_{t=1}^{N}(Y_t - F_{A,t})^2\right)^{\frac{1}{2}}}{\left(\prod_{t=1}^{N}(Y_t - F_{B,t})^2\right)^{\frac{1}{2}}} = \frac{\prod_{t=1}^{N}|Y_t - F_{A,t}|}{\prod_{t=1}^{N}|Y_t - F_{B,t}|} = \text{GMAE}$$

The only issue with calculating RGRMSE by Equation 2-24 is that if Y_t and $F_{B,t}$ are identical for a specific time period t, the measure will be undefined. Despite this problem, RGRMSE has been suggested as an appropriate measure for intermittent demand data (see for example Syntetos and Boylan, 2005), because in most of the cases the forecasts $(F_{B,t})$ are not integer (although averaging methods might produce integer forecasts) and therefore, the denominator is not zero. When all of the observations are zero, this shows that there has been no demand for the item in a long time and in reality no forecasts are made for such items. Therefore, such series can be excluded from the study.

2.4.3 Conclusions on Accuracy Measures

When comparing the performance of different forecasting methods, a range of accuracy measures should be used. This is because different measures are designed to assess different aspects. For example, MSE puts heavier penalties on higher errors while MAE is designed to lessen the effect of outliers. It has been found through forecasting competition studies such as the M-competition (Makridakis et al., 1982) and the M3-competition (Makridakis and Hibon, 2000) that the performance of different methods changes considerably depending on the accuracy measure being used. As a result, such studies have used several accuracy measures. As discussed previously, not all accuracy measures can be used for intermittent demand data.

MASE has been suggested for intermittent demand studies (Hyndman, 2006). Because it is based on in-sample MAE of the naïve method, there is no "division by zero" problem unless all of the observations are equal. However, MASE, as well as MAE, can sometimes be misleading when comparing forecasting methods.

The following example illustrates how the Zero Forecasting method (ZF) can show superiority to exponential smoothing when only MASE is concerned. The idea of using ZF as a benchmark comes from Teunter and Duncan (2009). They only used MAE and MSE, but we also include ME and MASE in our comparison. The data, the same as that presented by Hyndman (2006), are given in Table 2-2.

Table 2-2 Example data (Hyndman, 2006)

	In-sample	Out-of-sample
Actual Y_t	0 2 0 1 0 1 0 0 0 0 2 0 6 3 0 0 0 0 0 7 0 0 0 0	0 0 0 3 1 0 0 1 0 1 0 0
Naïve forecast $\hat{Y}_t$	0 2 0 1 0 1 0 0 0 0 2 0 6 3 0 0 0 0 0 7 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0

We have compared Exponential Smoothing (ES) with ZF. The results of ME, MASE, MAE, and MSE (average over both in-sample and out-of-sample errors) are given in Table 2-3.

Table 2-3 Comparing ES with ZF based on different accuracy measures

	ME	MASE	MAE	MSE
Exponential Smoothing $\alpha = 0.3$	0.1722	0.7114	1.1754	3.3458
Zero Forecast	0.8	0.4842	0.8	3.3143

The results of Table 2-3 suggest that ZF is the best method when MASE, MAE and MSE are used. Teunter and Duncan (2009) suggest that therefore these accuracy measures should not be used in this regard and inventory implication metrics should be used instead. However, we believe that these measures should be interpreted based on what they are designed to measure and, as a result, different measures should be used to compare forecasting methods. MASE and MAE are designed to measure the absolute errors while ME is designed to measure bias. Because intermittent data have many zero observations, when ZF is used on these observations, the error would be zero. Therefore, ZF produces better results on MASE and MAE than ES does. But looking at ME reveals that the forecasts are much more biased than forecasts based on ES.

Table 2-4 summarizes the absolute accuracy measures which will be used by this study.

Table 2-4 Accuracy measures for simulation and empirical studies

	Theory	Simulation	Empirical
MSE	✓	✓	✓
ME	✓	✓	✓
MASE		✓	✓
RGRMSE			✓
PB			✓

MSE is a widely used measure in the forecasting literature, is mathematically easy to handle, can be linked to a quadratic loss function and therefore it will be used for theoretical comparisons. MSE is a sensible measure for evaluating an individual time series and, although it is scale dependent and cannot be used for assessing a method's accuracy across multiple series, it can be used in simulation where all the series are theoretically generated. Keeping in mind the scale-dependency and sensitivity to outliers issues, we will also use MSE for empirical analysis.

ME can also be used for theoretical comparisons and, as discussed earlier, it does not have the scale-dependency problem. Therefore it will be used for both simulation and empirical analysis.

Based on the above mentioned properties, MASE will be used for simulation and empirical analysis. It should be mentioned that as both MAE and MASE are used to measure the same factor (absolute errors), we do not see any benefit in using both and only MASE will be calculated.

In addition to the above measures, two relative-to-another-method measures will be used for empirical analysis: PB and RGRMSE. The PB is based on comparing the MASEs of two methods and determines how often a method is better than another, but not by how much. RGRMSE is used to calculate the magnitude of improvement (in terms of MSE) of one method over another. Both of these measures have been used in an intermittent demand context (Syntetos and Boylan, 2005; Teunter and Duncan, 2009). They are insensitive to outliers and have been used in other studies, allowing comparative analyses to be undertaken.

2.5 Comparative Studies

A number of studies have been undertaken to compare methods for forecasting intermittent demand (Willemain et al., 1994; Johnston and Boylan, 1996; Eaves and Kingsman, 2004; Levén and Segerstedt, 2004; Willemain et al., 2004; Syntetos and Boylan, 2005). Three of the main comparison studies will be discussed here.

Eaves and Kingsman (2004) compare the performance of exponential smoothing (ES), Croston’s method, Syntetos-Boylan Approximation (SBA), a moving average (MA12) and a simple average method. Mean absolute deviation (MAD), root mean square error (RMSE) and MAPE have been used to compare the accuracy of forecasts obtained by each method, although it has been concluded that these measures are not ideal for slow-moving demands. It is also suggested to use stock-holdings as a measure instead of the above-mentioned conventional accuracy measures. The results show the superiority of SBA with regards to stock-holdings.

The study by Willemain et al. (2004) compares three methods of ES, Croston’s, and Willemain’s bootstrap based on the uniformity of observed lead-time demand (LTD) percentiles. It is found that Croston’s method provides a more accurate estimate of mean demand than ES, but the same is not true for forecasting the distribution of LTD. The results also show that bootstrapping is the most accurate method, especially for short lead times.

Another study by Syntetos and Boylan (2005) compares the accuracy of forecasts obtained by simple moving average (SMA13), SES, Croston's method, and SBA. ME, scaled mean error, RGRMSE, Percentage Better (PB), and Percentage Best (PBt) have been used and the results suggest that using different accuracy measures leads to different conclusions, agreeing with the findings of Eaves and Kingsman (2004). However, the SBA seems to be the most accurate method for faster intermittent demand (p values close to 1).

It is also argued by Teunter and Duncan (2009) that the conflicting results of comparative studies in the literature are due to use of inappropriate measures. Instead of comparing "per period forecast error" (measures such as MAD and MSE), they propose analyzing the effect of forecasting methods on inventory control parameters and also comparing the resulting average inventory and service levels. Boylan and Syntetos (2006) also highlight the distinction between these two approaches which they refer to as forecast-accuracy metrics and accuracy-implication metrics, respectively.

As discussed previously, because this research only focuses on forecasting and not on inventory control, the forecast-accuracy metrics suggested in section 2.4 will be used. The proposed method based on INARMA modeling of intermittent demand will be compared with Croston's method, SBA and SBJ. The reason for considering Croston's is its simplicity, popularity (for example, it is used in Forecast Pro), and its superiority over SES and SMA (Syntetos and Boylan, 2005). SBA has also been proved by a number of studies discussed above to perform reasonably well and therefore is included for comparison in this research. SBJ is included because it is based on Poisson demand arrivals. The reason for ignoring bootstrapping is that we have restricted our model to Poisson marginal distribution which makes it an inappropriate comparison considering the fact that there is no distributional assumption for bootstrapping.

2.6 INARMA Models

As discussed by Willemain et al. (1994), stochastic models of intermittent demand have assumed that the successive intervals between demands, successive demand sizes, and intervals and sizes are mutually independent. However, their sample data contradict this assumption (with some autocorrelations as high as -0.53 and 0.39). This emphasizes the necessity to use models that take into account autocorrelation.

Shenstone and Hyndman (2005) also suggest using integer autoregressive moving average (INARMA) models with Poisson marginal distribution or other time series models for counts for intermittent demand forecasting.

This research focuses on using a class of models called INARMA models to model intermittent demand. The forecasting method will then be the minimum mean square error (MMSE) of the model. We start from a stochastic model for intermittent demand and then build a method for forecasting which is optimal for the underlying model. It has also the advantage of enabling us to directly find the mean and variance of lead-time demand. The conditional mean of lead-time demand will be compared with the lead-time forecasts of the benchmark methods in chapters 8 and 9.

Table 2-5 summarizes the methods for intermittent demand forecasting based on their dependence, distributional, and stationarity assumptions. Although this research only focuses on INARMA models with Poisson marginal distribution, other discrete distributions such as negative binomial can be considered as a future line of study.

Table 2-5 The categorization of methods of intermittent demand forecasting based on their assumptions

Models		**Dependence structure**	**Distribution**
Croston's method		independent	Normal sizes of demand Geometric inter-arrival times
SBA		independent	Normal sizes of demand Geometric inter-arrival times
SBJ		independent	Geometric sizes of demand Negative Exponential or Erlang inter-arrival times
Bootstrapping:	Willemain et al. (2004)	Markov demand incidence	-
	Porras and Dekker (2007)	-	
INARMA modelling		autocorrelated demand	Poisson

2.7 Conclusions

The literature on forecasting intermittent demand has been reviewed in this chapter. Different definitions of intermittent demand have been reviewed. In this study, a series is called intermittent if it contains non-negative integer values with some values being zero (Shenstone and Hyndman, 2005). As explained, intermittence and lumpiness are two different concepts and should be distinguished. This study focuses on INARMA models with a Poisson marginal distribution. Although other marginal distributions could be used to allow for more demand size variability, in general these models are not designed for very lumpy demand. The two classification schemes for intermittent demand by Eaves and Kingsman (2004) and Syntetos et al. (2005) have been reviewed. As explained, the INARMA approach is universal and can be used for both intermittent and smooth demand but we will use the Syntetos et al. classification to find the best benchmark method for comparison.

Different intermittent demand forecasting methods have been reviewed. These methods are:

- Croston’s method and its variants
- Bootstrapping
- Causal methods

Although Croston's method is the most widely used approach for intermittent demand forecasting, it has been shown that it is biased (Syntetos and Boylan, 2001). Some modifications have been suggested to reduce its bias (e.g. SBA and SBJ) while some add even more bias (Modified Croston). The optimal demand model underlying Croston's method has not yet been established, although some studies have focused on finding such a model (Shenstone and Hyndman, 2005; Snyder and Beamont, 2007).

Different bootstrapping methods have been established in the literature to estimate the distribution of lead time demand. The parametric bootstrap by Snyder (2002) has the disadvantage of being based on the Normal distribution. The method of Willemain et al. (2004), however, is not based on any distributional assumptions. Porras and Dekker (2007) develop another procedure for identification of the lead time demand.

The studies on using causal models for intermittent demand forecasting have also been reviewed. Both bootstrapping methods and causal models seems promising but the literature on these methods is not yet well-developed. As will be discussed in chapter 10, incorporating causal factors into INARMA models can be pursued as a future line of study. The performance of INARMA models in forecasting the lead time distribution can also be compared to that of bootstrapping methods.

The fact that intermittent data include zero values makes some of the conventional accuracy measures inappropriate. Different accuracy measures that can be used for intermittent data are reviewed. Three absolute accuracy measures will be used in this research: ME, MSE, and MASE. For the empirical study, two relative-to-another-method accuracy measures will be added to the above-mentioned measures: PB and RGRMSE.

Using the INARMA procedure for intermittent demand forecasting has the advantage of utilizing a model-based forecasting method which takes into account the correlation between demands. Unlike bootstrapping, a marginal distribution has to be assumed. The Poisson distribution has been selected because it has been used in the IDF literature and also due to its interesting properties in INARMA models, to be reviewed in chapter 3.

As discussed in section 2.5, the results of INARMA method will be compared to Croston’s method, Syntetos-Boylan Approximation (SBA) and Shale-Boylan-Johnston (SBJ) method.

The next chapter introduces INARMA models and reviews their properties in detail.

Chapter 3 INTEGER AUTOREGRESSIVE MOVING AVERAGE MODELS

3.1 Introduction

Continuous-valued autoregressive integrated moving average (ARIMA) models, also known as Box-Jenkins models, developed by Box and Jenkins (1970) are used to model stationary processes under the assumption of Gaussianity, i.e. all the joint distributions of the time series are multivariate normal (Brockwell and Davis, 1996). However, this assumption is not valid for modelling count data, especially for low-frequency count data that cannot be suitably approximated by continuous models. Therefore, a number of models using different approaches have been proposed for integer-valued time series in the literature (e.g. Smith, 1979; Zeger, 1988; Zeger and Qaqish, 1988; Harvey and Fernandes, 1989).

These models are divided into two categories: observation-driven and parameter-driven. Observation-driven models specify a direct link between current and past observations, while parameter-driven models rely on a latent process connecting the observations (Jung and Tremayne, 2006a). As an example of parameter-driven models for time series of counts, Zeger and Qaqish (1988) proposed Poisson regression to model trend and seasonality explicitly and an unobserved stationary process (latent process) to model the autocorrelation.

One of the early observation-driven models for count data is the model suggested by Jacobs and Lewis (1978a; 1978b; 1983) called DARMA (discrete autoregressive moving average) models which will be discussed in section 3.2.

Another class of observation-driven models has been developed by McKenzie (1985) and later generalized by Al-Osh and Alzaid (1987) as integer-valued autoregressive (INAR) models for modelling series with correlated counting data. Examples of this kind of time series include the number of patients in a hospital at a specific point of time, the number of people in a queue waiting to receive a service at a certain moment (Silva and Oliveira, 2004), and the number of accidents in a firm each three months (McKenzie, 2003).

Table 3-1 briefly describes the most frequently used count models and summarizes their strengths and weaknesses (McKenzie, 2003; Rengifo, 2005).

This thesis is focused on integer autoregressive moving average (INARMA) models. As explained in chapter 2, the need for a model-based method for intermittent demand forecasting has motivated this research. The similarities that these models have with the conventional ARMA models are also an advantage. Because DARMA models are also based on ARMA models, we introduce them in section 3.2. The main reasons for excluding these models from our study are explained in this section.

This chapter mainly focuses on introducing the INARMA models and reviewing their statistical properties. Identification of these models, estimation of their parameters and INARMA forecasting will be discussed in subsequent chapters.

This chapter is structured as follows. In section 3.2, DARMA models are introduced.

INARMA models are reviewed in section 3.3. Summary of literature review and the conclusions are provided in sections 3.4 and 3.5.

Table 3-1 Review of count models in time series

Model	Description	Strengths and weaknesses
Markov chains	First, the transition probabilities between all the possible values that the count variable can take are defined. Then, the appropriate order of the time series is determined (e.g. see Raftery 1985, and Pegram 1980).	This method can only be reasonably used when the possible values that the observations can take are very limited. It is generally overparametrized and too limited in correlation structure.
Discrete Autoregressive Moving Average (DARMA) models	These models are structurally based of ARMA models and are probabilistic mixtures of discrete i.i.d. random variables with a marginal distribution (see section 3.2).	When the serial correlation is high, the data will be characterized by a series of runs of a single value. Therefore, they are rarely used. The main application is in the hydrological literature.
Integer Autoregressive Moving Average (INARMA) models	These models are a generalization of the linear ARMA models for count data (see section 3.3).	The models have the same serial correlation structure as ARMA models. The marginal distribution of the model is same as the distribution of the innovations if the latter is Poisson.
Regression models (or generalized linear models)	These are regression models for the special case where the dependent variable is a non-negative integer with a correction for autocorrelation (Zeger, 1988; Brännäs and Johansson, 1994).	These models are generally easy to construct and have the ovserdispersion property. An explicit joint density function cannot be obtained which restricts the class of possible predictors.
The hidden Markov models	These models are extension of the basic Markov chains models, in which various regimes characterizing the possible values of the mean are identified. It is then assumed that the transition from one regime to another is ruled by a Markov chain (MacDonald and Zucchini 1997).	There is no accepted way of determining the appropriate order for the Markov chain. There can be too many parameters to estimate, especially when the number of regimes is large.
State Space models	These models specify the conditional distribution of the series to depend on stochastic parameters that evolve according to a specified distribution. The parameters of such distributions are determined by some regressors (Harvey and Fernandes, 1989; Durbin and Koopman, 2000).	These models are very general and can be used in a wide range of applications. The behaviour of different components of the series can be modelled separately and then put together.

3.2 DARMA Models

This class of models was first introduced by Jacobs and Lewis (1978a) for a stationary sequence of dependent discrete random variables. Because we focus on INARMA models in this thesis, only a brief review of the first order discrete autoregressive model is provided in this section. The applications of DARMA models will then be discussed.

3.2.1 DAR(1) Model

The first order discrete autoregressive form, DAR(1), is given by:

$$Y_t = V_t Y_{t-1} + (1 - V_t) Z_t$$

Equation 3-1

where $\{V_t\}$ are i.i.d. binary random variables with $P(V_t = 1) = \alpha$ and $\{Z_t\}$ are i.i.d. discrete random variables with the probability mass function (hereafter called a "distribution" as by the authors) π[1] (Jacobs and Lewis, 1978a). Therefore, in this model, the current observation is either the previous observation, with probability α, or another, independent, sample from a specific distribution. This approach is similar to the Box-Jenkins approach except that a probabilistic mixture replaces the linear combination in the continuous-valued model and, as the authors mentioned, a realization of the process will generally contain many runs of a constant value. This can be a significant disadvantage of these models. It is obvious that the larger the value of α, the longer the runs.

It can be seen that, in the model of Equation 3-1, Y_{t-1} contains all information about the past. Because there is randomization between Y_{t-1} and Z_t, if Z_t is chosen, the memory of the process before time t is gone forever.

The autocorrelation function (ACF) of $\{Y_t\}$ is given by:

$$\rho_k = \alpha^k \qquad \text{for } k = 0,1, \dots$$

Equation 3-2

The conditional mean of Y_t given Y_{t-1} is linear in Y_{t-1} and the conditional variance is quadratic in Y_{t-1}. It also follows from Equation 3-1 that $\{Y_t\}$ is a Markov chain since:

$$P(Y_{t+1} = i | Y_1, \dots, Y_t) = P(Y_{t+1} = i | Y_t)$$

Equation 3-3

The transition matrix for this Markov chain is given by:

[1] $P(Z_t = i) = \pi(i) \qquad i = 0,1, \dots$

$$P(Y_{t+1} = i | Y_t = k) = \begin{cases} (1-\alpha)\pi(i) & \text{for } k \neq i \\ \alpha + (1-\alpha)\pi(i) & \text{for } k = i \end{cases}$$

Equation 3-4

It can be seen from Equation 3-1 that when $\alpha = 0$, the model is reduced to a stationary sequence of i.i.d. random variables with distribution π.

The mixed DARMA(p,q+1) model is built by adding the two autoregressive and moving average components as follows:

$$Y_t = U_t X_{t-S_t} + (1 - U_t) Z_{t-q-1}$$

Equation 3-5

$$Z_t = V_t Z_{t-A_t} + (1 - V_t) X_t$$

Equation 3-6

where $\{Z_t\}$ and $\{V_t\}$ are as before. $\{A_t\}$ are i.i.d. random variables defined on the set $k = \{1,2, \dots, p\}$ in such a way that $P(A_t = k) = \phi_k$. $\{S_t\}$ are i.i.d. random variables so that $P(S_t = k) = \theta_k$ for $k = \{0,1, \dots, q\}$ and $\{U_t\}$ are i.i.d. binary random variables with $P(U_t = 1) = \beta$.

3.2.2 Applications of DARMA Models

Applications of DARMA models can be found mainly in the hydrological literature. This is not surprising given the structure of the model, which represents dependence as runs (McKenzie, 2003). These applications are briefly discussed below.

In climatic analysis, these models were introduced by Buishand (1978). He proposes a binary discrete autoregressive moving average (B-DARMA) process to model daily rainfall sequences and finds that this model is promising in tropical and monsoonal areas.

Chang et al. (1984) use the same type of models (B-DARMA) in their study of daily precipitation by transforming the daily level of precipitation into a discrete variable based on its magnitude. They conclude that the statistical properties of the daily rainfall process can be preserved by DARMA models.

Salas et al. (2001) use DARMA models to simulate the return period and risk of extreme droughts. A series of wet and dry years is obtained from a continuous-valued hydrologic series, such as annual stream flows, and a method is presented for relating the autocorrelation functions of these two series. The analysis of 23 series of annual flows reveals that this relationship is applicable and reliable.

Ksenija (2006) employs a DARMA(1,1) model to describe the wet-dry day sequences in Split, on the middle Adriatic coast of Croatia. The results are compared to the DAR(1) model and reveal that, although both models underestimate dry spell runs, the DARMA(1,1) model provides a better fit to the empirical distribution both for short (one day) and long runs (more than 10 days). But, for short wet spells, the DAR(1) model estimates are closer to the observed frequencies of short spells in the months studied.

As previously mentioned, the main disadvantage of the DARMA models, which makes their application areas very limited, is that a realization of the process will generally contain many runs of a constant value. This is especially true when the serial correlation is high. Therefore we exclude these models and concentrate on another class of models for count data, called integer autoregressive moving average models, which overcome these problems.

3.3 INARMA Models

The integer-valued autoregressive models are equivalent to autoregressive models for Gaussian time series. By “equivalent” we mean that they share some similar properties, which will be discussed later in relevant subsections.

First, we introduce the first-order integer autoregressive, INAR(1), model. It has been shown that this model belongs to a more general class of models known as non-Gaussian conditional linear AR(1), CLAR(1), models (Grunwald et al., 2000). Other integer autoregressive, moving average and mixed models are then described.

Although many papers discuss the statistical aspects of INARMA models, fewer studies have been done regarding their practical application (Jung and Tremayne, 2003). Some applications of these models are reviewed at the end of this section.

3.3.1 INAR(1) Model

Before describing the INAR(1) model, we first introduce the meaning of the *binomial thinning* operation defined by Sueutel and van Harn (1979). Suppose Y is a non-negative integer-valued random variable. Then, for any $\alpha \in [0,1]$, the thinning operation "$\circ$" is defined by:

$$\alpha \circ Y = \sum_{i=1}^{Y} X_i$$

Equation 3-7

where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables, independent of Y, and with a constant probability that the variable will take the value of unity:

$$P(X_i = 1) = 1 - P(X_i = 0) = \alpha$$

Equation 3-8

From the above definition, some of the properties of the thinning operation can be obtained as follows:

(1) $0 \circ Y = 0$
(2) $1 \circ Y = Y$
(3) $\alpha \circ (\beta \circ Y) \stackrel{d}{=} (\alpha\beta) \circ Y$
(4) $E(\alpha \circ Y) = \alpha E(Y)$
(5) $E(\alpha \circ Y)^2 = \alpha^2 E(Y^2) + \alpha(1-\alpha)E(Y)$
(6) $\text{var}(\alpha \circ Y) = \alpha^2 \text{var}(Y) + \alpha(1-\alpha)E(Y)$

where $\stackrel{d}{=}$ stands for equal in distribution.

Now, the integer-valued first order autoregressive, INAR(1), model is defined by the Equation 3-9. A discrete time stochastic process, $\{Y_t\}$, is called an INAR(1) process if it satisfies the equation:

$$Y_t = \alpha \circ Y_{t-1} + Z_t$$

Equation 3-9

where $\alpha \in [0,1]$ and $\{Z_t\}$ is a sequence of i.i.d. non-negative integer-valued random variables, independent of Y_t with mean μ_Z and finite variance σ_Z^2. Z_t and Y_{t-1} are assumed to be stochastically independent for all points in time. The process obtained by the Equation 3-9 is stationary and it resembles the Gaussian AR(1) process except that it is nonlinear due to the thinning operation replacing the scalar multiplication in continuous models. It should be noted that, in Equation 3-9, subsequent thinning operations are performed independently of each other.

Equation 3-9 shows that, based on the definition of the thinning operation, unlike the DAR(1) model, the memory of an INAR(1) model decays exponentially (Al-Osh and Alzaid, 1987).

The two independence limitations we have assumed so far – independence of $\{X_i\}$ in the thinning operation, and independence of Z_t and Y_{t-1}– have been relaxed in a study by Brännäs and Hellström (2001).

It is worth mentioning that the probability α is assumed to be constant here. Alzaid and Al-Osh (1993) develop a model in which this probability of retaining an element is not constant. Also, Zheng et al. (2007) develop a random coefficient model where α_t are i.i.d. random variables that can take values in the interval $[0,1)$.

A realization of Y_t in an INAR(1) model of Equation 3-9 has two components: (i) the survivors of elements of the process at time $(t-1)$, Y_{t-1}, each with probability of survival α and (ii) the innovation term, Z_t, which represents new entrants to the system in the interval $(t-1, t]$.

The mean and variance of the process $\{Y_t\}$ are:

$$E(Y_t) = \alpha^t E(Y_0) + \mu_Z \sum_{j=0}^{t-1} \alpha^j$$

Equation 3-10

$$\mathrm{var}(Y_t) = \alpha^{2t}\mathrm{var}(Y_0) + (1-\alpha)\sum_{j=1}^{t} \alpha^{2j-1} E\left(Y_{t-j}\right) + \sigma_Z^2 \sum_{j=1}^{t} \alpha^{2(j-1)}$$

Equation 3-11

It is shown by Al-Osh and Alzaid (1987) that the autocorrelation function (ACF) of this process is given by $\rho_k = \alpha^k$ for $k = 1,2, \ldots$. This is identical to the ACF of a linear Gaussian AR(1) process. The only difference is that the ACF of the INAR(1) model is always positive.

The derivation of the first and second order unconditional moments of the INAR(1) process is straightforward:

$$E(Y_t) = \mu_Z/(1-\alpha)$$

Equation 3-12

$$\mathrm{var}(Y_t) = (\alpha\mu_Z + \sigma_Z^2)/(1-\alpha^2)$$

Equation 3-13

It can be seen that both the first order (regression function) and the second order conditional moments are linear in Y_{t-1}.

$$E(Y_t|Y_{t-1}) = \alpha Y_{t-1} + \mu_Z$$

Equation 3-14

$$\mathrm{var}(Y_t|Y_{t-1}) = \alpha(1-\alpha)Y_{t-1} + \sigma_Z^2$$

Equation 3-15

Another similarity between the model of Equation 3-9 and the Gaussian AR(1) is that the distribution of the innovation term (Z_t) plays the same role in determining the distribution of Y_t that the normal distribution of the shock term plays in the AR(1) model. In fact, Al-Osh and Alzaid (1987) argue that the distribution of Y_t is uniquely determined by the distribution of Z_t.

In Equation 3-9, $\{Z_t\}$ can have any non-negative discrete distribution. A natural first choice of interest for these variables is Poisson. Al-Osh and Alzaid (1987) show that if $Z_t \sim Po(\lambda)$, the marginal distribution of the process Y_t is also Poisson $Y_t \sim Po(\lambda/(1-\alpha))$. In this case the model is called PoINAR(1) (Jung and Tremayne, 2006b) (or PAR(1) as suggested, for example, by Freeland and McCabe (2004b)). Hence, the role of the Poisson distribution in the INAR(1) process is analogous to the role of the Gaussian distribution in the AR(1) process.

The INAR(1) process is a member of a class of models introduced by Grunwald et al. (2000). They suggest that nearly all the non-Gaussian AR(1) models are in fact a part of a class of conditional linear AR(1) models, CLAR(1).

If $\{Y_t\}, t = 0,1, \ldots$ is a time-homogeneous first-order Markov process on a sample space $\mathcal{Y} \in \mathbb{R}$, then it is said to have CLAR(1) structure if it satisfies:

$$m(Y_{t-1}) = \phi Y_{t-1} + \lambda$$

where $m(Y_{t-1}) = E(Y_t|Y_{t-1})$ and ϕ and λ are real numbers. Grunwald et al. (2000) show that stochastic properties of CLAR(1) models are similar to those of the Gaussian AR(1) model.

3.3.2 INAR(2) Model

To take into account higher order dependence in the data, higher order INAR models are developed. In this section, the second order model, INAR(2), will be reviewed briefly.

A discrete time stochastic process, $\{Y_t\}$ is called an INAR(2) process if it satisfies the equation:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t$$

Equation 3-16

with all the previously mentioned definitions, except for the thinning mechanism. There are two approaches regarding the binomial thinning mechanism.

The first approach is proposed by Alzaid and Al-Osh (1990) and the other by Du and Li (1991). The corresponding processes will henceforth be denoted by INAR(2)-AA and INAR(2)-DL, respectively, as in Jung and Tremayne (2006b).

In the INAR(2)-AA process, the random variables $\alpha_1 \circ Y_{t-2}$ and $\alpha_2 \circ Y_{t-2}$ which are elements of Y_{t-1} and Y_t, are connected in a powerful way. In fact, "$\alpha_1 \circ$" and "$\alpha_2 \circ$" and Y_{t-2} are dependent although they appear in different times. The vector $(\alpha_1 \circ Y_t, \alpha_2 \circ Y_t)$ given $Y_t = y_t$ is multinomial with parameters $(\alpha_1, \alpha_2, y_t)$. To understand this structure, consider the simulated process: at time t, Y_t is observed and from the previous time period we have Y_{t-1} and we have formed $\alpha_1 \circ Y_{t-1}$, and $\alpha_2 \circ Y_{t-1}$. Now, we form $U_t = \alpha_1 \circ Y_t$ and $V_t = \alpha_2 \circ Y_t$. Then we have $Y_{t+1} = U_t + V_{t-1} + Z_{t+1}$ and V_t is available to obtain Y_{t+2}.

This dependence results in two important consequences: (*i*) the process maintains its physical interpretation in terms of the counts evolving as a birth and death process, and (*ii*) a moving average structure is included into the process in such a way that the ACF of the INAR(2)-AA process mimics that of a Gaussian ARMA(2,1) process (Alzaid and Al-Osh, 1990).

The stationarity condition for the INAR(2)-AA process is $\alpha_1 + \alpha_2 < 1$ and, assuming $Z_t \sim Po(\lambda)$, the marginal distribution of the process Y_t is $Y_t \sim Po(\lambda/(1 - \alpha_1 - \alpha_2))$. However, the conditional mean function for this model is not linear, which shows that this process is not a member of the CLAR class.

The ACF of the INAR(2)-AA process is given by:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} \qquad \text{for } k \geq 2$$

Equation 3-17

where the starting values are $\rho_0 = 1$ and $\rho_1 = \alpha_1$ (Jung and Tremayne, 2006b).

The other specification of the INAR(2) model is that of Du and Li (1991). The concept of their model is closer to the Gaussian higher order AR models in which Y_t is obtained by a direct multiplication of the constants α_1 and α_2 to Y_{t-1} and Y_{t-2}, independent of all previous stochastic structures. It means that at time t we have already observed Y_{t-1} and Y_{t-2} and the thinning operations are applied independently of the previous period which result in: $Y_t = (\alpha_1 \circ Y_{t-1}) + (\alpha_2 \circ Y_{t-2}) + Z_t$. Du and Li (1991) show that the unconditional mean of the process is again $\mu_Z/(1 - \alpha_1 - \alpha_2)$ and the stationarity condition remains the same. However, they show that the correlation properties of their model are identical to those of the Gaussian AR(2) model which is one of the main differences that distinguishes INAR(2)-DL from INAR(2)-AA. An additional difference is that the conditional mean of the Y_t in INAR(2)-DL is given by:

$$E(Y_t | Y_{t-1}, Y_{t-2}, \ldots) = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu_Z$$

Equation 3-18

which obviously is linear, while that of the INAR(2)-AA process is nonlinear. Therefore, the INAR(2)-DL model has a CLAR(2) structure. It should also be noted that, even with Poisson innovations, the marginal distribution of Y_t is not Poisson, which is in contrast with the INAR(2)-AA model.

Because the specification by Du and Li (1991) is similar to the conventional AR(2) model and we do not use the physical interpretation of the INAR models (as a birth and death process which is maintained in the specification by Alzaid and Al-Osh (1990)), we use the Du and Li approach in this study.

3.3.3 INAR(p) Model

Realizations of some counting process $\{Y_t\}$ might be attributed not only to its immediate predecessors $\{Y_{t-1}\}$ and $\{Y_{t-2}\}$ as in INAR(2), but also to previous realizations of the process, $\{Y_{t-j}\}_{j=3}^{p}$. The pth order integer-valued autoregressive, INAR(p), process is defined by Alzaid and Al-Osh (1990) as follows:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \cdots + \alpha_p \circ Y_{t-p} + Z_t$$

Equation 3-19

with all the previously mentioned definitions. $\{\alpha_j\}$ are non-negative constants such that the process remains stationary and $\alpha_1, \ldots, \alpha_{p-1} \in [0,1]$ and $\alpha_p \in (0,1]$. The stationarity condition for the INAR(p) process is that the roots of the equation $x^p - \alpha_1 x^{p-1} - \cdots - \alpha_{p-1} x - \alpha_p = 0$ lie inside the unit circle (Alzaid and Al-Osh, 1990).

Like the INAR(2), there are two approaches concerning the binomial thinning mechanisms. The first model is that of Alzaid and Al-Osh (1990) which will be denoted by INAR(p)-AA and the other is the model of Du and Li (1991), known as INAR(p)-DL. The idea behind these approaches is the same as that explained for INAR(2). The INAR(p)-AA process shares the same correlation properties with the Gaussian ARMA(p,p-1) process (Alzaid and Al-Osh, 1990), while the INAR(p)-DL mimics the AR(p) process (Du and Li, 1991).

The unconditional first moment of the INAR(p)-AA process is given by:

$$E(Y_t) = \sum_{i=1}^{p} \alpha_i E(Y_{t-i}) + \mu_Z$$

Equation 3-20

The conditional moments of Y_t are given by:

$$E(Y_t|Y_{t-1},\dots,Y_{t-p}) = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \mu_Z$$

$$\text{var}(Y_t|Y_{t-1},\dots,Y_{t-p}) = \sum_{i=1}^{p} \alpha_i(1-\alpha_i)Y_{t-i} + \sigma_Z^2$$

Assuming $Z_t \sim Po(\lambda)$, the marginal distribution of the process Y_t would be $Y_t \sim Po(\lambda/(1-\sum_{i=1}^{p}\alpha_i))$.

The autocovariance at lag k of the INAR(p)-AA process satisfies the equation:

$$\gamma_k = \sum_{i=1}^{p} \alpha_i \gamma_{k-i} + \sum_{i=k+1}^{p} \mu_{k-i,\alpha_i} + \delta_k(0)\sigma_Z^2$$

Equation 3-21

where $\mu_{k-i,\alpha_i} \equiv \text{cov}(Y_{n-k+i}, \alpha_i \circ Y_n) - \alpha_i \gamma_l$ and $\delta_k(0) = 1$ if $k = 0$ and zero otherwise. The autocovariance of the INAR(p)-AA has the same form of that of the Gaussian ARMA(p,p-1).

As for the INAR(2)-DL, when the approach of Du and Li (1991) is taken, the marginal distribution of Y_t is not the same as the distribution of innovations. The autocovariance function of the INAR(p)-DL satisfies:

$$\gamma_k = \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p}$$

Equation 3-22

Therefore, the ACF of this process is found from equations of the form:

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p}$$

Equation 3-23

which implies that the correlation structures of INAR(p) and AR(p) processes are the same. This makes these models similar to the standard AR(p) models not only in form, but also in stationarity conditions and correlation structure.

3.3.4 INMA(1) Model

Having introduced integer autoregressive models for count series, Al-Osh and Alzaid (1988) then developed a class of models for integer-valued moving average (INMA) processes. In INMA models, a stationary sequence of random variables $\{Y_t\}$ is formed from a sequence $\{Z_t\}$ of i.i.d. random variables which are non-negative and also integer-valued. The first order model, which we are going to describe in this section, is the case in which adjacent members of the sequence are correlated. A process $\{Y_t\}$ is called an INMA(1) process if it satisfies the equation:

$$Y_t = \beta \circ Z_{t-1} + Z_t$$

Equation 3-24

where $\beta \in [0,1]$ and $\{Z_t\}$ are as before and the thinning operation is defined via:

$$\beta \circ Z = \sum_{i=1}^{Z} X_i$$

Equation 3-25

where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables, independent of Y and satisfying:

$$P(X_i = 1) = 1 - P(X_i = 0) = \beta$$

Equation 3-26

The INMA(1) model defined by Equation 3-24 is similar to the Gaussian MA(1) process except that scalar multiplication is replaced by the thinning operation. Jung and Tremayne (2006a) present a physical interpretation of this model as follows. If we consider Y_t as the number of particles in a well-defined space at time point t, it can be assumed that this number is made of two components: (i) particles entering during $(t-1, t]$, and (ii) survivors of those who entered the space during $(t-2, t-1]$. Therefore, the thinning at time t, is applied to only immigrants at time $t-1$, not all particles in space, as in an INAR(1) process. Examples of this process include the number of patients staying in a hospital or the number of customers in a department store (Al-Osh and Alzaid, 1988).

It can be inferred from the Equation 3-24 that each element stays in the system no longer than two periods. This is in contrast to the INAR(1) process in which there is no limit on the survival of elements in the system.

The unconditional first and second moments of the INMA(1) process are:

$$E(Y_t) = (1+\beta)\mu_Z$$

Equation 3-27

$$\mathrm{var}(Y_t) = \beta(1-\beta)\mu_Z + (1+\beta^2)\sigma_Z^2$$

Equation 3-28

It is shown by Al-Osh and Alzaid (1988) that the autocorrelation function (ACF) of this process is given by:

$$\rho_k^{\text{INMA (1)}} = \begin{cases} \dfrac{\beta\sigma_Z^2}{\beta(1-\beta)\mu_Z + (1+\beta^2)\sigma_Z^2} & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases}$$

Equation 3-29

which is analogous (but not identical) to that of the Gaussian MA(1) process, where $Y_t = \theta Z_{t-1} + Z_t$ and the ACF is given by:

$$\rho_k^{\text{MA (1)}} = \begin{cases} 1 & \text{for } k = 0 \\ \dfrac{\theta}{1+\theta^2} & \text{for } k = \pm 1 \\ 0 & \text{for } |k| > 1 \end{cases}$$

Equation 3-30

Another property of the INMA(1) process which is similar to MA(1) is that if $\beta = 0$, the sequence $\{Y_t\}$ becomes a sequence of i.i.d. random variables with the distribution of Z_t. Also, if $\beta = 1$ the process will have the highest ρ_1 which again agrees with the MA(1) process.

As for the INAR(1) process, a natural candidate for the marginal distribution of an INMA(1) process is the Poisson distribution. It is shown by Al-Osh and Alzaid (1988) that assuming $Z_t \sim Po(\lambda)$, the marginal distribution of the process Y_t would be $Y_t \sim Po(\lambda(1+\beta))$. This process is referred to as a PoINMA(1) process (Jung and Tremayne, 2006a).

3.3.5 INMA(2) Model

The second-order moving average process is an extension of the INMA(1) process introduced in the previous section. The model is given by:

$$Y_t = \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2} + Z_t$$

Equation 3-31

where both parameters β_1 and β_2 lie in the interval $[0,1]$. The individual thinning operations $\beta_j \circ Z_{t-j}$ for $j = 1,2$ follow the Equation 3-25 and it is assumed that they perform independently of each other. Two approaches arise regarding the thinning mechanisms, as in higher order autoregressive models.

One approach is proposed by Al-Osh and Alzaid (1988), assuming dependence between the thinnings of terms $\beta_1 \circ Z_t$ and $\beta_2 \circ Z_t$ (similar to that of INAR(2)-AA). The unconditional expected value and variance of such a process, henceforth called INMA(2)-AA, are given by:

$$E(Y_t) = \mu_Z(1 + \beta_1 + \beta_2)$$

Equation 3-32

$$\mathrm{var}(Y_t) = \mu_Z \sum_{j=1}^{2} \beta_j\left(1 - \beta_j\right) + \sigma_Z^2 \sum_{j=0}^{2} \beta_j^2$$

Equation 3-33

with $\beta_0 = 1$. The ACF of the INMA(2)-AA process is:

$$\rho_k = \begin{cases} \dfrac{\sum_{j=0}^{2-k}[\beta_j(\beta_k - \beta_{k+j})\mu_Z + \beta_j\beta_{k+j}\sigma_Z^2]}{\mu_Z\sum_{j=1}^{2}\beta_j(1-\beta_j) + \sigma_Z^2\sum_{j=0}^{2}\beta_j^2} & \text{for } k = 1,2 \\ 0 & \text{for } k > 2 \end{cases}$$

Equation 3-34

It can be seen that the cut-off property of the INMA(2)-AA process is the same as that of the Gaussian MA(2) process.

The other approach concerning the thinning operation in an INMA(2) process is introduced by McKenzie (1988). In an INMA(2)-MK process, it is assumed that the individual thinning operations $\beta_j \circ Z_{t-j}$ for $j = 1,2$ are performed independently not only from each other, but also from corresponding operations at previous times in

Equation 3-31. The unconditional moments of this process with Poisson innovations are the same as INMA(2)-AA, $E(Y_t) = \mathrm{var}(Y_t) = \lambda(1 + \beta_1 + \beta_2)$, and the ACF, which is again the same as that of the Gaussian MA(2) process, is given by:

$$\rho_k = \begin{cases} \dfrac{\sum_{j=0}^{2-k} \beta_j \beta_{k+j}}{1 + \beta_1 + \beta_2} & \text{for } k = 1,2 \\ 0 & \text{for } k > 2 \end{cases}$$

Equation 3-35

Unlike INAR(2)-DL, for an INMA(2)-MK process, if $Z_t \sim Po(\lambda)$, then $Y_t \sim Po(\lambda(1 + \beta_1 + \beta_2))$ which is the same as INMA(2)-AA. In this PhD thesis, we adopt the approach by McKenzie (1988) because his model is more similar to the classic MA(q) model (Brännäs and Hall, 2001).

3.3.6 INMA(q) Model

The qth order integer moving average model, introduced by Al-Osh and Alzaid (1988) and McKenzie (1988) is defined by:

$$Y_t = \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2} + \cdots + \beta_q \circ Z_{t-q} + Z_t$$

Equation 3-36

where $\{Z_t\}$ is defined as before and the parameters $\beta_1, \ldots, \beta_{q-1} \in [0,1]$ and $\beta_q \in (0,1]$.

Using the properties of the thinning operation, it is shown by Brännäs and Hall (2001) that:

$$E(Y_t) = \mu_Z(1 + \sum_{j=1}^{q} \beta_j)$$

Equation 3-37

$$\mathrm{var}(Y_t) = \sigma_Z^2 + \sum_{j=1}^{q} [\sigma_Z^2 \beta_j^2 + \mu_Z \beta_j (1 - \beta_j)]$$

Equation 3-38

As for an INMA(2) process, there are two approaches based on the thinning mechanisms. The ACF of the INMA(q)-AA process is given by:

$$\rho_k = \begin{cases} \dfrac{\sum_{j=0}^{q-k}[\beta_j(\beta_k - \beta_{k+j})\mu_Z + \beta_j\beta_{k+j}\sigma_Z^2]}{\mu_Z\sum_{j=1}^{q}\beta_j(1-\beta_j) + \sigma_Z^2\sum_{j=0}^{q}\beta_j^2} & \text{for } k = 1, \dots, q \\ 0 & \text{for } k > q \end{cases}$$

Equation 3-39

On the other hand, the ACF of the INMA(q)-MK process is:

$$\rho_k = \begin{cases} \dfrac{\sum_{j=0}^{q-k}\beta_j\beta_{k+j}}{\sum_{j=0}^{q}\beta_j} & \text{for } k = 1, \dots, q \\ 0 & \text{for } k > q \end{cases}$$

Equation 3-40

It can be seen that the autocorrelation function of an INMA(q) process is analogous to than of the classical MA(q) process. The difference is that all autocorrelations are positive (Brännäs and Hall, 2001).

3.3.7 INARMA(1,1) Model

Having introduced INAR and INMA processes, Alzaid and Al-Osh (1990) suggested that these two processes can be mixed in a manner similar to that of the standard ARMA processes to provide the mixed integer autoregressive moving average class of models. There are two approaches regarding the modelling of this kind of processes.

The first approach was introduced by McKenzie (1988) for INARMA processes with Poisson marginal distributions. He suggests the mixed process should be constructed by coupling the two AR and MA processes and a common innovation process. According to this viewpoint, the AR component of the INARMA(1,1) process is given by:

$$Y_t = \alpha \circ Y_{t-1} + Z_t$$

Equation 3-41

and the MA component is:

$$W_t = Y_{t-1} + \beta \circ Z_t$$

Equation 3-42

where all the thinning operations are independent, $\alpha, \beta \in [0,1]$ and $\{Z_t\}$ is a sequence of i.i.d. Poisson variables.

The second approach suggested by Neal and Rao (2007) is what we follow because it is similar to the Gaussian ARMA process. A discrete time stochastic process, $\{Y_t\}$, is called an INARMA(1,1) process if it satisfies the equation:

$$Y_t = \alpha \circ Y_{t-1} + Z_t + \beta \circ Z_{t-1}$$

Equation 3-43

where $\alpha, \beta \in [0,1]$ and $\{Z_t\}$ is a sequence of i.i.d. non-negative integer-valued random variables, independent of Y_t with mean μ_Z and finite variance σ_Z^2. Here, the two thinning operations are independent of each other and also of the corresponding operations at previous times, and are defined as follows:

$$\alpha \circ Y = \sum_{i=1}^{Y} X_i$$

Equation 3-44

$$\beta \circ Z = \sum_{j=1}^{Z} X_j$$

Equation 3-45

To ensure the stationarity and invertibility of the above INARMA(1,1) process given by Equation 3-43, the two conditions of $\alpha < 1$ and $\beta < 1$ must hold.

The unconditional first and second moments of this process are:

$$E(Y_t) = \left(\frac{1+\beta}{1-\alpha}\right)\mu_Z$$

Equation 3-46

$$\text{var}(Y_t) = \frac{1}{1-\alpha^2}\left[(\alpha + \alpha\beta + \beta - \beta^2)\mu_Z + (1 + \beta^2 + 2\alpha\beta)\sigma_Z^2\right]$$

Equation 3-47

The autocorrelation function (ACF) of this process is given by (see Appendix 3.A for the proof):

$$\rho_k^{\text{INARMA }(1,1)} = \begin{cases} \dfrac{(\alpha^2+\alpha^2\beta+\alpha\beta-\alpha\beta^2)\mu_Z+(\alpha+\alpha\beta^2+\alpha^2\beta+\beta)\sigma_Z^2}{(\alpha+\alpha\beta+\beta-\beta^2)\mu_Z+(1+\beta^2+2\alpha\beta)\sigma_Z^2} & \text{for } k=1 \\ \alpha\rho_{k-1} & \text{for } k>1 \end{cases}$$

Equation 3-48

It can be seen that the ACF of an INARMA(1,1) dies exponentially, which is analogous to the ACF of the Gaussian ARMA(1,1) which is as follows (for $\mu_Z = 0$ and $\sigma_Z^2 = 1$):

$$\rho_k^{\text{ARMA }(1,1)} = \begin{cases} \dfrac{\alpha+\beta+\alpha^2\beta+\alpha\beta^2}{1+\beta^2+2\alpha\beta} & \text{for } k=1 \\ \alpha\rho_{k-1} & \text{for } k>1 \end{cases}$$

Equation 3-49

3.3.8 INARMA(*p*,*q*) Model

The INARMA(*p*,*q*) process is given by the following difference equation:

$$Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i \circ Z_{t-i}$$

Equation 3-50

where $\alpha_i, \beta_i \in [0,1]$ and $\{Z_t\}$ is as before and thinning operations are performed independently of each other and also of the corresponding operations at previous times.

The stationarity conditions of this process are the same as those of an INAR(*p*) process: to ensure that the above process is stationary, it is required that $(\alpha_1, \alpha_2, \dots, \alpha_p)$ are such that the roots of the *p*-order polynomial $x^p - \alpha_1 x^{p-1} - \cdots - \alpha_{p-1}x - \alpha_p = 0$ lie inside the unit circle.

Neal and Rao (2007) discuss the invertibility conditions for an INARMA(*p*,*q*) process for the moving average parameters $(\beta_1, \beta_2, \dots, \beta_q)$. They assume that these conditions are the same as the those of an MA(*q*) process. However, they have not provided any proof in this regard and left it as an open question to investigate if this condition is sufficient for an INARMA(*p*,*q*) process to be invertible.

3.3.8.1 First and Second Unconditional Moments

As mentioned before, the stochastic properties, including the autocorrelation function, of the general INARMA(p,q) process have not been found in the literature. Therefore, to answer the first research question "How can the appropriate integer autoregressive moving average (INARMA) model be identified for a time series of counts?", here we investigate these properties.

Obtaining the first unconditional moment of the INARMA process of Equation 3-50 is straightforward. It is given by:

$$E(Y_t) = \left(\frac{1 + \sum_{i=1}^{q} \beta_i}{1 - \sum_{i=1}^{p} \alpha_i}\right) \mu_Z$$

Equation 3-51

However, the derivation of the second unconditional moment is more challenging. We have found the unconditional variance of the INARMA process of Equation 3-50 as follows (see Appendix 3.B for the proof):

$$\begin{aligned}\mathrm{var}(Y_t) &= \frac{\mu_Z}{1 - \sum_{i=1}^{p} \alpha_i^2}\left[\frac{1 + \sum_{i=1}^{q} \beta_i}{1 - \sum_{i=1}^{p} \alpha_i}\sum_{i=1}^{p} \alpha_i(1 - \alpha_i) + \sum_{i=1}^{q} \beta_i(1 - \beta_i)\right] \\ &+ \frac{\sigma_Z^2}{1 - \sum_{i=1}^{p} \alpha_i^2}\left[1 + \sum_{i=1}^{q} \beta_i^2 + 2\sum_{i=1}^{\min(p,q)} \alpha_i\beta_i\right] \\ &+ \frac{2\sum_{j=1}^{p-1}\sum_{i=1}^{p-j} \alpha_i\alpha_{i+j}\gamma_j + 2\sum_{i=1}^{\min(p,q)}\sum_{j=i+1}^{q} \alpha_i\beta_j\gamma_{j-i}^{YZ}}{1 - \sum_{i=1}^{p} \alpha_i^2}\end{aligned}$$

Equation 3-52

where γ_k^{YZ} is the cross-covariance function derived in Appendix 3.C and γ_k is the autocovariance at lag k which can be expressed in terms of γ_0 (or $\mathrm{var}(Y_t)$) from the equations obtained in the next section.

It can be seen that if $q = 0$, for an INAR(p) process, the unconditional variance would be:

$$\mathrm{var}(Y_t) = \frac{1}{1 - \sum_{i=1}^{p} \alpha_i^2}\left\{\left[\frac{\sum_{i=1}^{p} \alpha_i(1 - \alpha_i)}{1 - \sum_{j=1}^{p} \alpha_i}\right]\mu_Z + \sigma_Z^2 + 2\sum_{j=1}^{p-1}\sum_{i=1}^{p-j} \alpha_i\alpha_{i+j}\gamma_j\right\}$$

Equation 3-53

In the above equation, γ_k can be expressed in terms of γ_0 based on the Yule-Walker equations of Equation 4-12. For example, for an INAR(2) process we have:

$$\gamma_1 = \alpha_1\gamma_0 + \alpha_2\gamma_1$$

and as a result:

$$\gamma_1 = \frac{\alpha_1}{1-\alpha_2}\gamma_0$$

The unconditional variance of an INAR(2) process can then be found from Equation 3-53 to be:

$$\mathrm{var}(Y_t) = \frac{(\alpha_1 - \alpha_1^2 + \alpha_2 - 2\alpha_2^2 - \alpha_1\alpha_2 + \alpha_1^2\alpha_2 + \alpha_2^3)\mu_Z + (1 - \alpha_1 - 2\alpha_2 + \alpha_1\alpha_2 + \alpha_2^2)\sigma_Z^2}{1 - \alpha_1 - \alpha_1^2 + \alpha_1^3 - 2\alpha_2 + 2\alpha_2^3 - \alpha_2^4 + \alpha_1\alpha_2 + \alpha_1\alpha_2^2 - \alpha_1\alpha_2^3 + \alpha_1^2\alpha_2}$$

It can be seen from Equation 3-53 that the unconditional mean and variance of a PoINAR(p) process are not equal when the distribution of the innovations is Poisson. This has been also confirmed in the literature (Bu and McCabe, 2008).

Also, when $p = 0$, for an INMA(q) process, the variance would be:

$$\mathrm{var}(Y_t) = \left[\sum_{i=1}^{q}\beta_i(1-\beta_i)\right]\mu_Z + \left[1 + \sum_{i=1}^{q}\beta_i^2\right]\sigma_Z^2$$

Equation 3-54

which agrees with the result found by Brännäs and Hall (2001) (Equation 3-38). Equation 3-52 is the first new result found in this PhD study.

3.3.8.2 Autocorrelation Function (ACF)

The next step is finding the autocorrelation function of an INARMA(p,q) process. In order to do so, first, we need to find the covariance of this process. The covariance of INARMA(p,q) at lag k is:

$$\gamma_k = \mathrm{cov}(Y_t, Y_{t-k})$$

According to the relation between k and p and q, there are four cases:

- $k \le p,\ k \le q$
- $k \le p,\ k > q$
- $k > p,\ k \le q$
- $k > p,\ k > q$

Each of these cases will be considered in sequence.

1. If $k \le p,\ k \le q$ (k can be only equal to either p or q if $p \ne q$)

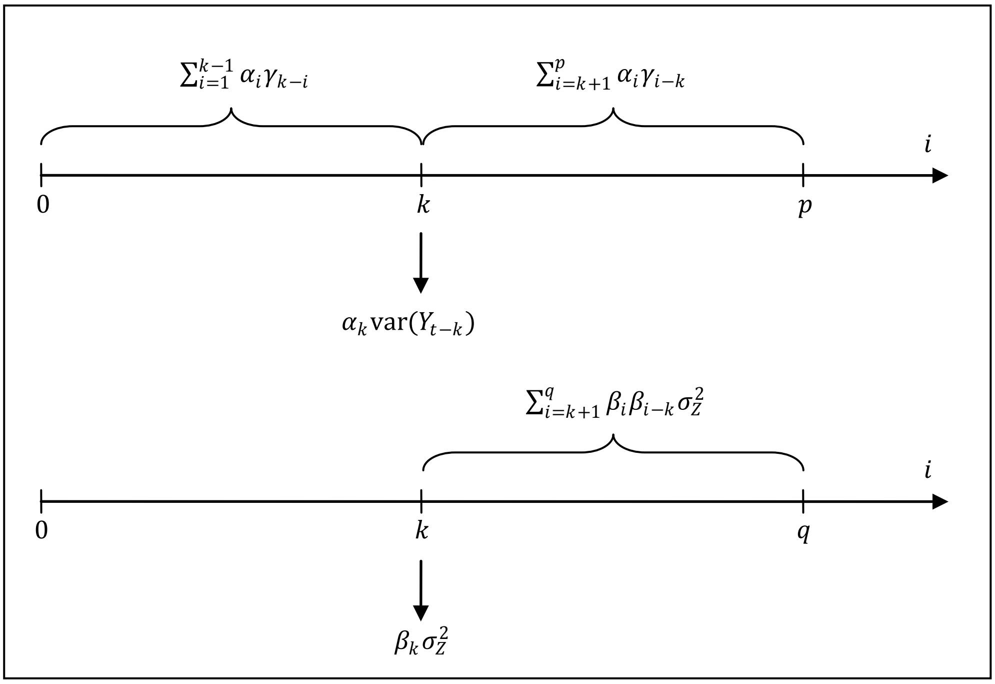

Figure 3-1 The covariance at lag k, γ_k, when $k \le p, k \le q$

$$
\begin{aligned}
\gamma_k &= \mathrm{cov}(Y_t, Y_{t-k}) \\
&= \mathrm{cov}\left[\left(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}\right), Y_{t-k}\right] \\
&= \mathrm{cov}[(\alpha_1 \circ Y_{t-1}), Y_{t-k}] + \cdots + \mathrm{cov}\left[\left(\alpha_p \circ Y_{t-p}\right), Y_{t-k}\right] \\
&\quad + \mathrm{cov}[(\beta_k \circ Z_{t-k}), Y_{t-k}] + \mathrm{cov}[(\beta_{k+1} \circ Z_{t-k-1}), Y_{t-k}] + \cdots \\
&\quad + \mathrm{cov}\left[\left(\beta_{q-1} \circ Z_{t-q+1}\right), Y_{t-k}\right] + \mathrm{cov}\left[\left(\beta_q \circ Z_{t-q}\right), Y_{t-k}\right]
\end{aligned}
$$

$$
\begin{aligned}
\gamma_k &= \alpha_1 \gamma_{k-1} + \cdots + \alpha_{k-1}\gamma_1 + \alpha_k \mathrm{var}(Y_{t-k}) + \alpha_{k+1}\gamma_1 + \cdots + \alpha_p \gamma_{p-k} \\
&\quad + \beta_k \gamma_0^{YZ} + \beta_{k+1}\gamma_1^{YZ} + \cdots + \beta_{q-1}\gamma_{(q-1)-k}^{YZ} + \beta_q \gamma_{q-k}^{YZ}
\end{aligned}
$$

where γ_k^{YZ} is the cross-covariance between Y and Z at lag k (see Appendix 3.C). Then, the above equation can be written as:

$$\gamma_k = \sum_{i=1}^{k-1} \alpha_i \gamma_{k-i} + \alpha_k \mathrm{var}(Y_{t-k}) + \sum_{i=k+1}^{p} \alpha_i \gamma_{i-k} + \beta_k \sigma_Z^2 + \sum_{i=k+1}^{q} \beta_i \gamma_{i-k}^{YZ}$$

Note that if $2k \leq p$, there will be a γ_k in $\sum_{i=k+1}^{p} \alpha_i \gamma_{i-k}$ which has to be considered. This is because when $i = 2k$, we have $\alpha_{2k} \gamma_{2k-k} = \alpha_{2k} \gamma_k$.

2. If $k \leq p,\ k > q$

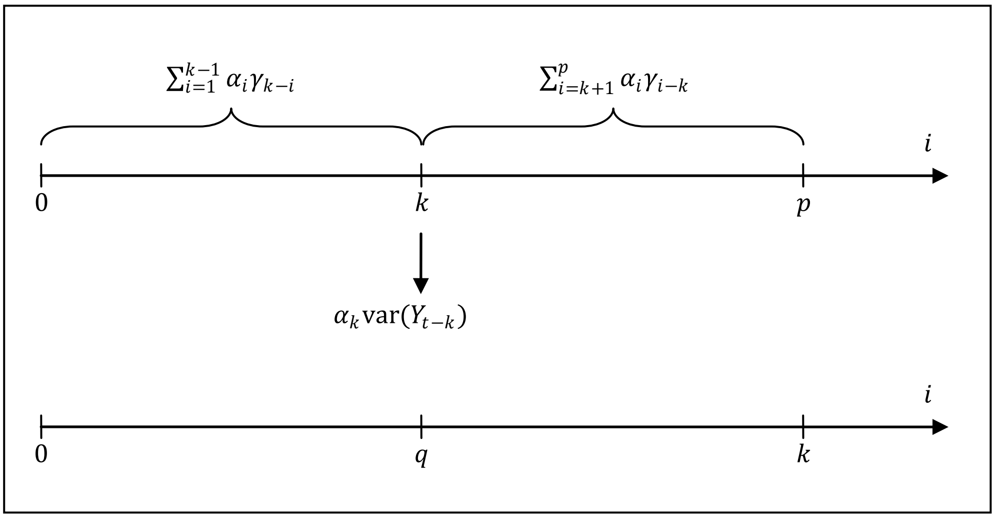

Figure 3-2 The covariance at lag at lag k , γ_k , when $k \leq p, k > q$

$$\begin{aligned}
\gamma_k &= \mathrm{cov}(Y_t, Y_{t-k}) \\
&= \mathrm{cov}\big[\big(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}\big), Y_{t-k}\big] \\
&= \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_{k-1} \gamma_1 + \alpha_k \mathrm{var}(Y_{t-k}) + \alpha_{k+1} \gamma_1 + \alpha_{k+2} \gamma_2 + \cdots + \alpha_p \gamma_{p-k}
\end{aligned}$$

Therefore, for the second case, the autocovariance at lag k can be obtained from:

$$\gamma_k = \sum_{i=1}^{k-1} \alpha_i \gamma_{k-i} + \alpha_k \mathrm{var}(Y_{t-k}) + \sum_{i=k+1}^{p} \alpha_i \gamma_{i-k}$$

Again, if $2k \leq p$, there will be a γ_k in $\sum_{i=k+1}^{p} \alpha_i \gamma_{i-k}$ which has to be considered.

3. If $k > p,\ k \le q$

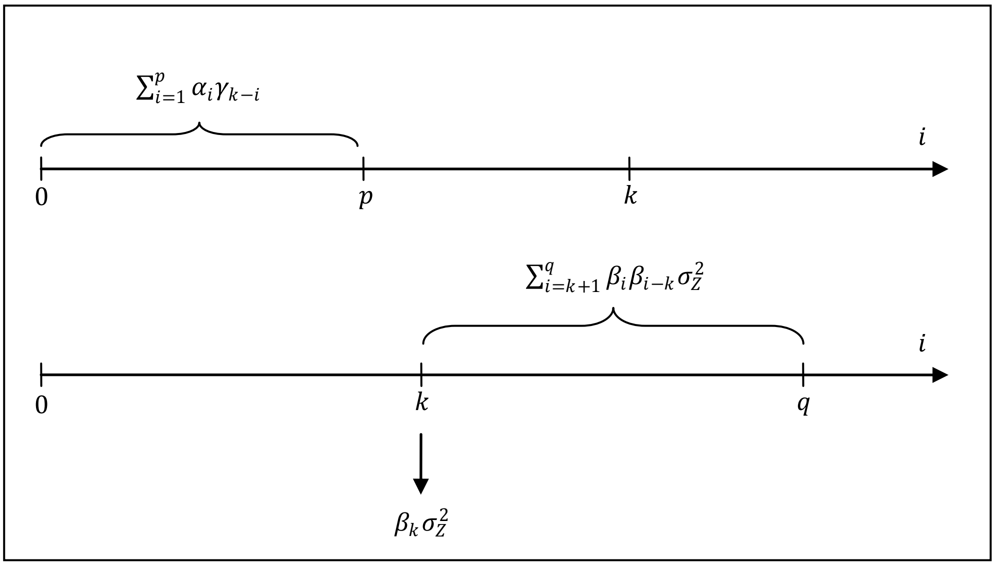

Figure 3-3 The covariance at lag k , γ_k , when $k > p, k \le q$

$$
\begin{aligned}
\gamma_k &= \mathrm{cov}(Y_t, Y_{t-k}) \\
&= \mathrm{cov}\big[\big(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}\big), Y_{t-k}\big] \\
&= \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_p \gamma_{k-p} + \beta_k \gamma_0^{YZ} + \beta_{k+1} \gamma_1^{YZ} + \cdots + \beta_{q-1} \gamma_{(q-1)-k}^{YZ} + \beta_q \gamma_{q-k}^{YZ}
\end{aligned}
$$

Therefore, the autocovariance at lag k for the third case can be obtained from:

$$
\gamma_k = \sum_{i=1}^{p} \alpha_i \gamma_{k-i} + \beta_k \sigma_Z^2 + \sum_{i=k+1}^{q} \beta_i \gamma_{i-k}^{YZ}
$$

4. If $k > p,\ k > q$

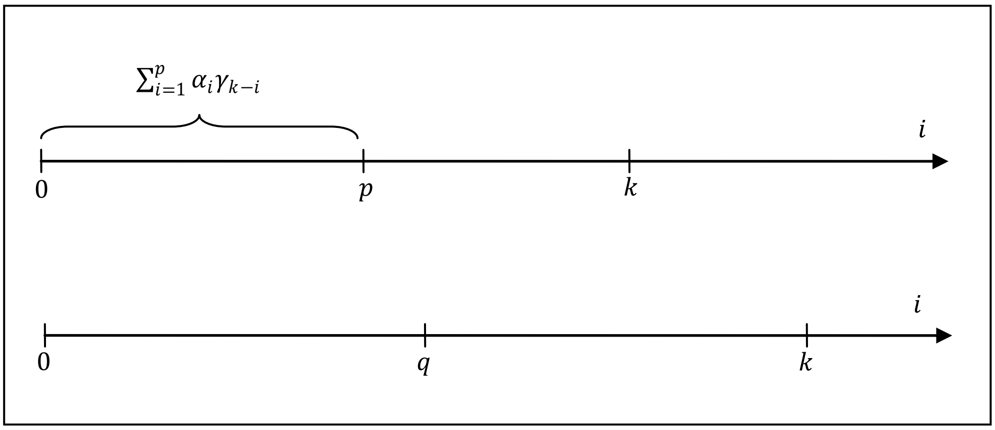

Figure 3-4 The covariance at lag k , γ_k , when $k > p, k > q$

$$\begin{aligned}\gamma_k &= \mathrm{cov}(Y_t, Y_{t-k}) \\ &= \mathrm{cov}\left[\left(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}\right), Y_{t-k}\right] \\ &= \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} = \sum_{i=1}^{p} \alpha_i \gamma_{k-i}\end{aligned}$$

Finally, the autocovariance at lag k for the fourth case is given by:

$$\gamma_k = \sum_{i=1}^{p} \alpha_i \gamma_{k-i}$$

Therefore, we can write the autocorrelation function of an INARMA(p,q) process as follows:

$$\rho_k = \begin{cases} \begin{cases} \dfrac{\sum_{i=1}^{k-1}\alpha_i\gamma_{k-i} + \alpha_k var + \sum_{i=k+1}^{2k-1}\alpha_i\gamma_{i-k} + \sum_{i=2k+1}^{p}\alpha_i\gamma_{i-k} + \beta_k\sigma_Z^2 + \sum_{i=k+1}^{q}\beta_i\gamma_{i-k}^{YZ}}{(1-\alpha_{2k})var} & \text{for } 2k \le p \\ \dfrac{\sum_{i=1}^{k-1}\alpha_i\gamma_{k-i} + \alpha_k var + \sum_{i=k+1}^{p}\alpha_i\gamma_{i-k} + \beta_k\sigma_Z^2 + \sum_{i=k+1}^{q}\beta_i\gamma_{i-k}^{YZ}}{var} & \text{for } 2k > p \end{cases} & k \le p, k \le q \\ \begin{cases} \dfrac{\sum_{i=1}^{k-1}\alpha_i\gamma_{k-i} + \alpha_k var + \sum_{i=k+1}^{2k-1}\alpha_i\gamma_{i-k} + \sum_{i=2k+1}^{p}\alpha_i\gamma_{i-k}}{(1-\alpha_{2k})var} & \text{for } 2k \le p \\ \dfrac{\sum_{i=1}^{k-1}\alpha_i\gamma_{k-i} + \alpha_k var + \sum_{i=k+1}^{p}\alpha_i\gamma_{i-k}}{var} & \text{for } 2k > p \end{cases} & k \le p, k > q \end{cases}$$

Equation 3-55

where var is the second unconditional moment of the process given by the Equation 3-52 and γ_k^{YZ} is the cross-covariance given by the Equation 3.C-1 (see Appendix 3.C). The other two cases of $k > p, k \le q$ and $k > p, k > q$ are special cases of the above expressions.

The above equation can be simply written as:

$$\rho_k = \begin{cases} \dfrac{\alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} + \beta_k\gamma_0^{YZ} + \beta_{k+1}\gamma_1^{YZ} + \cdots + \beta_q\gamma_{q-k}^{YZ}}{\gamma_0} & k \le q \\ \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p} & k > q \end{cases}$$

Equation 3-56

Equation 3-56 is another new result of this research. Identification of the order of an INARMA(p,q) process requires both the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the process. The structure of the PACF of an INARMA(p,q) process will be discussed in chapter 4.

In order to test the Equation 3-56, we check if it results in the correct ACF for INAR(p) and INMA(q) processes. When $q = 0$, i.e. for an INAR(p) process, it can be seen that the ACF based on the Equation 3-56 will be:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p}$$

which agrees with the result found by Du and Li (1991).

When $p = 0$, i.e. for an INMA(q) process, it can be seen that the ACF based on Equation 3-56 will be:

$$\rho_k = \begin{cases} \dfrac{\beta_k \sigma_Z^2 + \sum_{i=k+1}^{q} \beta_i \gamma_{i-k}^{YZ}}{\gamma_0} & k \leq q \\ 0 & k > q \end{cases}$$

where $var = \left[\sum_{j=1}^{q} \beta_j (1 - \beta_j)\right] \mu_Z + \left[1 + \sum_{j=1}^{q} \beta_j^2\right] \sigma_Z^2$. For an INMA($q$) process with Poisson marginal distribution ($\mu_Z = \sigma_Z^2 = \lambda$), the variance is given by:

$$var = \left[1 + \sum_{j=1}^{q} \beta_j\right] \lambda$$

It can be seen from Equation 3.C-1 that for an INMA(q) process $\gamma_k^{YZ} = \beta_k \lambda$ for $0 \leq k \leq q$, so we have:

$$\rho_k = \begin{cases} \dfrac{\beta_k + \sum_{i=k+1}^{q} \beta_i \beta_{i-k}}{1 + \sum_{j=1}^{q} \beta_j} & k \leq q \\ 0 & k > q \end{cases}$$

which agrees with the result in the literature (Brännäs and Hall, 2001).

3.3.9 Applications of INARMA Models

Applications of INAR processes in the medical sciences can be found in, for example, Franke and Seligmann (1993) and Cardinal et al. (1999); and applications to economics in, for example, Böckenholt (1999), Berglund and Brännäs (2001), Brännäs and Hellström (2001), Rudholm (2001) and Freeland and McCabe (2004b).

Brännäs (1995) studies the consequences and required adaptations when explanatory variables are included in the INAR(1) model. He obtains new conditional least squares (CLS) and generalized method of moments estimators for the model containing explanatory variables and applies the INAR(1) model with Poisson marginal distribution to the number of Swedish mechanical paper and pulp mills during 1921-1981.

INAR(1) models also have applications in inventory control. Aggoun et al. (1997) use an INAR(1) inventory model for perishable items, i.e. each item in the stock perishes in a given time period with an unknown probability. The sequence of these probabilities is assumed to be a homogeneous Markov chain and the paper finds the conditional probability distribution of this sequence and estimates the transition probabilities of the Markov chain. However, the model is not applied to real-world data.

Cardinal et al. (1999) represent infectious disease incidence time series by INAR models. They state that real-valued time series models have been used in the analysis of infectious disease surveillance data, but argue that these models are not suitable in some cases such as the analysis of a rare disease. Meningococcal infection is considered as a rare disease in their study and the integer-valued INAR(5) model is fitted to the data set.

Böckenholt (1999) introduces the application of INAR models in investigating regularity and predictability of purchase behaviour over time. He uses a PoINAR(1) model for the analysis of longitudinal purchase data because there is a notion that purchase behaviour of nondurable goods is well-described by a Poisson process. The population of consumers is then divided to an unknown number of mutually exclusive and exhaustive segments, and within each a PoINAR(1) process is used to model the counts. The mixed PoINAR(1) model is finally applied to a powder detergent purchase data set of about 5000 households.

Another straightforward application of INAR models can be found in a paper by Berglund and Brännäs (2001) in which they study the entry and exit of plants in Swedish municipalities as an INAR(1) process. In their model, they incorporate the variables affecting survival and entry and employ generalized method of moments for estimation.

Brännäs and Hellström (2001) apply the INAR(1) model to the number of Swedish mechanical paper and pulp mills. They consider the number of firms in a region at a certain time to be equal to the number of firms surviving from the previous time period plus the number of new firms. They assume that the survival of a firm depends on the survival of other firms (dependent exits) and there is dependence between the survival of a firm and entry of a new firm (dependent entry and exit). The results for the set of data show that the correlation between exits is not significant, but that of entry and exit is significant. In another Swedish application, Rudholm (2001) analyses the factors affecting entry into the Swedish pharmaceuticals market using an INAR(1) model.

Brännäs et al. (2002) find another application of INAR models in forecasting hotel guest nights which conventionally is based on economic demand models, pure time series analytical models, or on a mixture of them. They suggest that the daily number of guest nights for a specific hotel follows an INAR(1) process, and then proceed by cross-sectional and temporal aggregation of the model. The former means aggregation over more than two hotels which yields an INAR(1) model, and the latter means aggregation over time that results in an INARMA(1,1) model, later simplified as an INMA(1) model.

Karlis (2002) introduces an INAR(1) model with a general mixed Poisson distribution for the innovation term which allows for overdispersion. The model is then applied to the number of forest fires in Greece in a two-month period (daily observations).

Blundell et al. (2002) apply a Linear Feedback Model (LFM), which is derived from the INAR process, to the panel data of Hall et al. (1986). In their model, the technological output of a firm is a function of the corresponding R&D investment in current and previous periods, some unknown technology parameters, and the firm-specific propensity to patent. As another application, Gourieroux and Jasiak (2004) use an INAR(1) model to update premiums in car insurance and compare it to the standard negative binomial approach.

Freeland and McCabe (2004b) use a PoINAR model for a monthly count data set of claimants for wage loss benefit. They propose a method for producing coherent forecasts based on the conditional median rather than the conventional conditional mean. It is also argued that when the counts are low, the median should be accompanied by estimates of the probabilities associated with the point masses of the h-step-ahead conditional distribution. A method for calculating confidence intervals for these probabilities is also presented.

Quddus (2008) uses a PoINAR(1) model for analysis of traffic accidents in Great Britain. The results of his study show that an ARIMA model performs better than INAR(1) for geographically and temporally aggregated time series. However, as expected, the reverse is true for disaggregated low count time series (see section 3.3.10 for aggregation in INARMA models).

It can be seen that INARMA models have been generally used for time series of counts (counts of objects, events, or individuals). Although this application area is based on the direct physical interpretation of the INARMA models, as suggested in the literature (McKenzie, 2003), this should not restrict these models to only such applications. This research is an attempt to use INARMA models beyond their direct interpretation to model intermittent demand.

3.3.10 Aggregation in INARMA Models

Time series aggregation is a widely discussed subject for continuous-valued time series. It goes back over 50 years (Quenouille, 1958) and since then many papers have considered different aspects of aggregation in continuous-valued time series (see for example: Amemiya and Wu (1972), Brewer (1973), Harvey and Pierse (1984), Nijman and Palm (1990), Drost and Nijman (1993), Marcellino (1999), Teles and Wei (2002), and Man (2004)).

Three types of aggregation have been identified in the literature which can be classified as:

a. temporal aggregation
b. cross-sectional aggregation
c. over a forecast horizon aggregation

Temporal aggregation, also called flow scheme, refers to aggregation in which a low frequency time series (e.g. annual) is achieved from a high frequency time series (e.g. quarterly or monthly). It means that the low frequency variable is the sum of *k* consecutive periods of the high frequency variable. For example, the annual observations are the sum of the monthly observations every twelve periods. For Gaussian models, it has been proved that the aggregation of an AR(*p*) process produces an ARMA(*p*,*q*) process where $q < p$ (Amemiya and Wu, 1972). It means that not only the data is aggregated, but also the model is aggregated. In other words, the aggregate model can be inferred from the disaggregate model, i.e. the parameters of the aggregate model can be estimated based on the disaggregate data. The practical application of this is that as soon as new disaggregate observations are available the aggregate parameters can be inferred. This is a very useful tool for situations where decisions are taken, say, annually, but information is available, for instance, monthly. Here there is no need to wait till the end of the year to update the parameters of the aggregate series, but the annual model can be updated as soon as monthly observations become available.

The second class of aggregation is cross-sectional or contemporaneous aggregation. This scheme of aggregation is conducted through individuals rather than time. For example, in demand forecasting of many products with a short demand history, similar products are grouped in a product family and the demand forecast is built for the family rather than individuals, which may produce more reliable forecasts than the forecasts for individual items.

Finally, over a forecast horizon aggregation refers to the case in which a forecast is needed for some periods of time ahead. For example, in demand forecasting in a supply chain, when there is a lead time between ordering by a manufacturer and receiving the order from a supplier, the demand over that lead time has to be forecasted in order to prevent shortage during the lead time period (see for example Lee et al., 2000).

Although many papers examine different types of aggregation in continuous-valued time series, the same is not true for time series of counts. It might be due to the complicated probabilistic structure of these models that only a few papers address this issue. Brännäs et al. (2002) first studied temporal and cross-sectional aggregation of an INAR(1) process. In the following sections we review the work by Brännäs et al. (2002) for the first class of aggregation and provide the results of aggregation of an INAR(1) process over l periods. The third class of aggregation will be discussed in detail for an INARMA(p,q) process in chapter 6 when we study the lead time aggregation of this process.

3.3.10.1 Overlapping Temporal Aggregation

In this section we look at the properties of aggregating the INAR(1) process over two periods with overlap, so that if $Y_t = \alpha \circ Y_{t-1} + Z_t$, the new sequence $\{Y_{t,2}\}$ is defined as:

$$Y_{t,2} = Y_t + Y_{t+1}$$

$$Y_{t+1,2} = Y_{t+1} + Y_{t+2}$$

$$\vdots$$

$$Y_{t+n,2} = Y_{t+n} + Y_{t+n+1}$$

The aggregated innovations are also presented by:

$$Z_{t,2} = Z_t + Z_{t+1}$$

$$Z_{t+1,2} = Z_{t+1} + Z_{t+2}$$

$$\vdots$$

$$Z_{t+n,2} = Z_{t+n} + Z_{t+n+1}$$

Now, if we add each two subsequent terms in the sequence $\{Y_t\}$ with overlap, we have:

$$Y_{t+1,2} = Y_{t+1} + Y_{t+2} = \alpha \circ Y_t + Z_{t+1} + \alpha \circ Y_{t+1} + Z_{t+2}$$

Equation 3-57

But, we know that $\alpha \circ Y_t + \alpha \circ Y_{t+1} \overset{d}{=} \alpha \circ (Y_t + Y_{t+1})$. Now, from the Equation 3-57 we have:

$$Y_{t+1,2} = \alpha \circ (Y_t + Y_{t+1}) + (Z_{t+1} + Z_{t+2}) = \alpha \circ Y_{t,2} + Z_{t+1,2}$$

Therefore, $Y_{t,2}$ is also an INAR(1) process.

The same argument can be used to show that the overlapping aggregation of an INAR(1) process over $(l+1)$ periods also results in an INAR(1) process.

$$Y_{t,l+1} = Y_t + \cdots + Y_{t+l}$$

$$Z_{t,l+1} = Z_t + \cdots + Z_{t+l}$$

$$Y_{t+1,l+1} = Y_{t+1} + \cdots + Y_{t+l+1} = \alpha \circ Y_{t,l+1} + Z_{t+1,l+1}$$

which is an INAR(1) process with the same autoregressive parameter and, for a Po-INAR(1) model, the innovation parameter is $(l+1)\lambda$.

3.3.10.2 Non-overlapping Temporal Aggregation

In this section, we investigate the non-overlapping temporal aggregation of an INAR(1) model. Brännäs et al. (2002) suggest that aggregation of INAR(1) models produces an INARMA(1,1) model which is analogous to AR(1) models.

Let Y_t for $t = 0,1,\ldots$ be a high frequency or disaggregate time series observed at time t which follows an INAR(1) process ($Y_t = \alpha \circ Y_{t-1} + Z_t$). The low frequency or aggregate variable is assumed to be available only every kth period $(k, 2k, 3k, \ldots)$, where k is an integer value larger that one. Note that the aggregated series is not overlapped. For example, one year does not overlap with the next one. Therefore, to show the aggregate series we define another time scale T that runs in kt periods. So that $t = 0,1,\ldots$ while $T = 0, k, 2k, \ldots$. We show the aggregate series by Y_T^*. This is shown in Figure 3-5 for aggregating a monthly time series to an annual one.

$$Y_{t+1} = \alpha \circ Y_t + Z_{t+1}$$

$$Y_{t+2} = \alpha \circ Y_{t+1} + Z_{t+2} = \alpha \circ [\alpha \circ Y_t + Z_{t+1}] + Z_{t+2} = \alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + Z_{t+2}$$

Repeated use of the above equation yields:

$$Y_{t+L} = \alpha^L \circ Y_t + \sum_{i=1}^{L} \alpha^{L-i} \circ Z_{t+i}$$

Equation 3-58

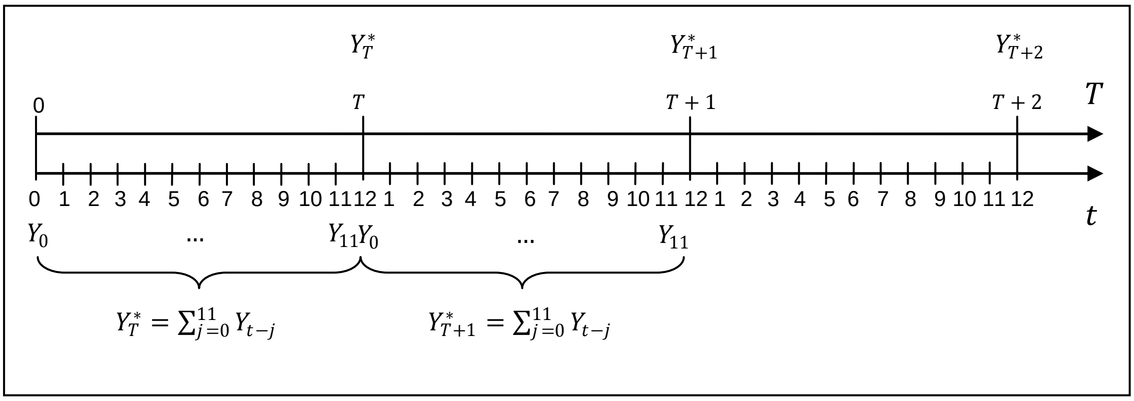

Figure 3-5 Non-overlapping temporal aggregation for $k = 12$

For example, for aggregation over two periods $(Y_1 + Y_2, Y_3 + Y_4, \dots)$ the aggregate series $(\dots, X_{\tau-1}, X_\tau, X_{\tau+1}, \dots)$ can be written as:

$$\begin{aligned}
X_\tau &= Y_{t-1} + Y_t = \alpha \circ Y_{t-2} + Z_{t-1} + \alpha \circ Y_{t-1} + Z_t \\
&= \alpha \circ (\alpha \circ Y_{t-3} + Z_{t-2}) + Z_{t-1} + \alpha \circ (\alpha \circ Y_{t-2} + Z_{t-1}) + Z_t \\
&= \alpha^2 \circ Y_{t-3} + \alpha \circ Z_{t-2} + Z_{t-1} + \alpha^2 \circ Y_{t-2} + \alpha \circ Z_{t-1} + Z_t \\
&= \alpha^2 \circ (Y_{t-3} + Y_{t-2}) + (Z_{t-1} + Z_t) + \alpha \circ (Z_{t-1} + Z_{t-2}) \\
&= \alpha^2 \circ X_{\tau-1} + \xi_\tau + \alpha \circ \xi_{\tau-1}
\end{aligned}$$

But we have:

$$\begin{aligned}
X_{\tau-1} &= Y_{t-3} + Y_{t-2} = \alpha \circ Y_{t-4} + Z_{t-3} + \alpha \circ Y_{t-3} + Z_{t-2} \\
&= \alpha \circ (\alpha \circ Y_{t-5} + Z_{t-4}) + Z_{t-3} + \alpha \circ (\alpha \circ Y_{t-4} + Z_{t-3}) + Z_{t-2} \\
&= \alpha^2 \circ Y_{t-5} + \alpha \circ Z_{t-4} + Z_{t-3} + \alpha^2 \circ Y_{t-4} + \alpha \circ Z_{t-3} + Z_{t-2} \\
&= \alpha^2 \circ (Y_{t-5} + Y_{t-4}) + (Z_{t-2} + Z_{t-3}) + \alpha \circ (Z_{t-3} + Z_{t-4}) \\
&= \alpha^2 \circ X_{\tau-2} + \xi_{\tau-1} + \alpha \circ \xi_{\tau-2}
\end{aligned}$$

Brännäs et al. (2002) state that this is an INARMA(1,1) model where the INAR(1) parameter (α^2) is the square of the original INAR(1) parameter, α, and the INMA(1) parameter is α. It is also mentioned that in the INMA-part, ξ_τ and $\xi_{\tau-1}$ are correlated which makes the INMA-part unconventional, but with interpretable parameters.

They also mention that the ξ_τ term of the X_τ expression is different from the ξ_τ term of the $X_{\tau+1}$ expression because the temporal aggregation is non-overlapping. These should be distinguished by different indexing as $\xi_\tau^{t,t-1}$ and $\xi_\tau^{t+1,t}$ respectively.

$$X_{\tau-1} = Y_{t-3} + Y_{t-2} = \alpha^2 \circ X_{\tau-2} + \xi_{\tau-1}^{t-3,t-2} + \alpha \circ \xi_{\tau-2}^{t-4,t-3}$$

$$X_{\tau} = Y_{t-1} + Y_{t} = \alpha^2 \circ X_{\tau-1} + \xi_{\tau}^{t-1,t} + \alpha \circ \xi_{\tau-1}^{t-2,t-1}$$

$$X_{\tau+1} = Y_{t+1} + Y_{t+2} = \alpha^2 \circ X_{\tau} + \xi_{\tau+1}^{t+1,t+2} + \alpha \circ \xi_{\tau}^{t,t+1}$$

This is shown in Figure 3-6.

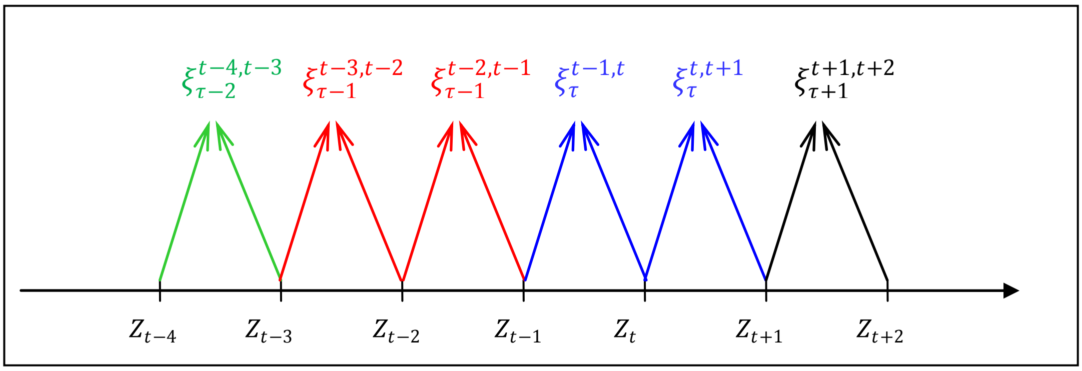

Figure 3-6 The correlation between the INMA parts of the non-overlapping temporal aggregation of INAR(1) process over two periods

If it is assumed that $\upsilon_t = (Z_{t-1} + Z_t) + \alpha \circ (Z_{t-1} + Z_{t-2})$ and $\omega_\tau = \xi_\tau + \alpha \circ \xi_{\tau-1}$, Brännäs et al. (2002) find the first and second moments of υ_t and ω_t for a PoINAR(1) process. They also analyze covariance properties, as summarized below.

$$E(\upsilon_t) = 2\lambda + 2\alpha\lambda$$

$$\begin{aligned}\mathrm{var}(\upsilon_t) &= \mathrm{var}(Z_{t-1} + Z_t) + \mathrm{var}[\alpha \circ (Z_{t-1} + Z_{t-2})] + 2\mathrm{cov}(Z_{t-1}, \alpha \circ Z_{t-1}) \\ &= 2\lambda + 2\alpha^2\lambda + 2\alpha(1-\alpha)\lambda + 2\alpha\lambda = 2\lambda + 4\alpha\lambda\end{aligned}$$

$$\begin{aligned}&\mathrm{cov}(\upsilon_t, \upsilon_{t-2}) \\ &\quad = \mathrm{cov}\{[(Z_{t-1} + Z_t) + \alpha \circ (Z_{t-1} + Z_{t-2})], [(Z_{t-3} + Z_{t-2}) + \alpha \\ &\quad \circ (Z_{t-3} + Z_{t-4})]\} \\ &\quad = \mathrm{cov}(\alpha \circ Z_{t-2}, Z_{t-2}) = \alpha\lambda\end{aligned}$$

and for ω_τ we have

$$E(\omega_\tau) = 2\lambda + 2\alpha\lambda$$

$$\begin{aligned}\mathrm{var}(\omega_\tau) &= \mathrm{var}(\xi_\tau) + \mathrm{var}(\alpha \circ \xi_{\tau-1}) + 2\mathrm{cov}(\xi_\tau, \alpha \circ \xi_{\tau-1}) \\ &= 2\lambda + 2\alpha^2\lambda + 2\alpha(1-\alpha)\lambda + 2\alpha\lambda = 2\lambda + 4\alpha\lambda\end{aligned}$$

$$\mathrm{cov}(\omega_\tau, \omega_{\tau-1}) = \mathrm{cov}[(\xi_\tau + \alpha \circ \xi_{\tau-1}), (\xi_{\tau-1} + \alpha \circ \xi_{\tau-2})] = \mathrm{cov}(\alpha \circ \xi_{\tau-1}, \xi_{\tau-1})$$
$$= \mathrm{cov}[\alpha \circ (Z_{t-1} + Z_{t-2}), (Z_{t-2} + Z_{t-3})] = \alpha\lambda$$

which can be seen to be equal to the first and second moments and covariance of v_t. However, having two versions of ξ_τ, $(\xi_\tau^{t-1,t}, \xi_\tau^{t,t+1})$, will result in:

$$\mathrm{cov}(\xi_\tau^{t-1,t}, \xi_\tau^{t,t+1}) = \mathrm{cov}[(Z_{t-1} + Z_t), (Z_t + Z_{t+1})] = \mathrm{var}(Z_t) = \lambda$$

$$\mathrm{corr}(\xi_\tau^{t-1,t}, \xi_\tau^{t,t+1}) = \frac{\lambda}{2\lambda} = \frac{1}{2}$$

Therefore, the covariance and correlation are halved due to this dependency.

Brännäs et al. (2002) also suggest that the non-overlapping temporal aggregation of an INAR(1) process over more than two periods (h periods) result in an INARMA(1,1) with the form of:

$$X_\tau = \alpha^h \circ X_{\tau-1} + \xi_\tau + \left(\sum_{i=1}^{h-1} \alpha^i\right) \circ \xi_{\tau-1}$$

Equation 3-59

and assume that $\sum_{i=1}^{h-1} \alpha^i = \beta$.

The $\left(\sum_{i=1}^{h-1} \alpha^i\right) \circ \xi_{\tau-1}$ term in the Equation 3-59 comes from the expression $\alpha \circ \xi_{\tau-1} + \alpha^2 \circ \xi_{\tau-1} + \cdots + \alpha^{h-1} \circ \xi_{\tau-1}$ again with the same assumptions for ξ_τs as discussed before (see Appendix A of Brännäs et al., 2002).

However, here another problem arises regarding the properties of the thinning operation. The problem is that $\alpha \circ X + \alpha \circ Y = \alpha \circ (X + Y)$, but $\alpha \circ X + \beta \circ X \neq (\alpha + \beta) \circ X$. Therefore, finding the result of non-overlapping temporal aggregation over h periods also remains unsolved. This line of research will not be taken forward in this thesis, but remains for further research (see chapter 10).

3.3.10.3 Cross-sectional Aggregation

As mentioned earlier, cross-sectional aggregation is over individual series. We start again by the simplest case of adding two INAR(1) processes. Suppose that X_t and W_t

are two PoINAR(1) processes:

$$X_t = \alpha \circ X_{t-1} + u_t$$

$$W_t = \beta \circ W_{t-1} + \upsilon_t$$

where u_t and υ_t are uncorrelated for all t, and $E(u_t) = \text{var}(u_t) = \lambda_u$ and $E(\upsilon_t) = \text{var}(\upsilon_t) = \lambda_\upsilon$.

From the above equations it can be found that:

$$X_{t-1} = \alpha \circ X_{t-2} + u_{t-1}$$
$$\Rightarrow \beta \circ X_{t-1} = \beta \circ (\alpha \circ X_{t-2} + u_{t-1}) = \alpha\beta \circ X_{t-2} + \beta \circ u_{t-1}$$

$$W_{t-1} = \beta \circ W_{t-2} + \upsilon_{t-1}$$
$$\Rightarrow \alpha \circ W_{t-1} = \alpha \circ (\beta \circ W_{t-2} + \upsilon_{t-1}) = \alpha\beta \circ W_{t-2} + \alpha \circ \upsilon_{t-1}$$

so we have:

$$X_t - (\beta \circ X_{t-1}) = (\alpha \circ X_{t-1} + u_t) - (\alpha\beta \circ X_{t-2} + \beta \circ u_{t-1})$$

$$W_t - (\alpha \circ W_{t-1}) = (\beta \circ W_{t-1} + \upsilon_t) - (\alpha\beta \circ W_{t-2} + \alpha \circ \upsilon_{t-1})$$

therefore,

$$X_t = \beta \circ X_{t-1} + \alpha \circ X_{t-1} + u_t - \alpha\beta \circ X_{t-2} - \beta \circ u_{t-1}$$

$$W_t = \alpha \circ W_{t-1} + \beta \circ W_{t-1} + \upsilon_t - \alpha\beta \circ W_{t-2} - \alpha \circ \upsilon_{t-1}$$

The cross-sectionally aggregated Y_t is found as:

$$\begin{aligned}
Y_t &= X_t + W_t \\
&= \alpha \circ X_{t-1} + \alpha \circ W_{t-1} + \beta \circ X_{t-1} + \beta \circ W_{t-1} - \alpha\beta \circ X_{t-2} - \alpha\beta \circ W_{t-2} \\
&\quad + u_t - \beta \circ u_{t-1} + \upsilon_t - \alpha \circ \upsilon_{t-1} \\
&= \alpha \circ Y_{t-1} + \beta \circ Y_{t-1} - \alpha\beta \circ Y_{t-2} + u_t - \beta \circ u_{t-1} + \upsilon_t - \alpha \circ \upsilon_{t-1}
\end{aligned}$$

The conditional expected value of this process is given by:

$$\begin{aligned}
E(Y_t | Y_{t-1}, Y_{t-2}) &= \alpha Y_{t-1} + \beta Y_{t-1} - \alpha\beta Y_{t-2} + \lambda_u - \beta\lambda_u + \lambda_\upsilon - \alpha\lambda_\upsilon \\
&= (\alpha + \beta) Y_{t-1} + (-\alpha\beta) Y_{t-2} + \lambda_u - \beta\lambda_u + \lambda_\upsilon - \alpha\lambda_\upsilon
\end{aligned}$$

Which is equal to the expected value of an INARMA(2,1) process of:

$$Y_t = \theta_1 \circ Y_{t-1} + \theta_2 \circ Y_{t-2} + Z_t + \phi Z_{t-1}$$

where $\theta_1 = (\alpha + \beta)$ and $\theta_2 = -\alpha\beta$ and ϕ is to be found. The AR part is a CLAR(2) in which:

$$E(Y_t | Y_{t-1}, Y_{t-2}) = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu_Z$$

Now, we want to find what is the MA part in sum of two INAR(1) processes. First, we show that sum of the conditional expected values of two INMA(1) processes is equal to the conditional expected value of an INMA(1) process. Then we obtain the parameters of the aggregated process (Y_t) in terms of the parameters of X_t and W_t.

Sum of two INMA(1) processes

Consider the case of adding two Gaussian MA(1) processes:

$$X_t = u_t + \alpha u_{t-1}$$

$$W_t = v_t + \beta v_{t-1}$$

The aggregated series is shown by Y_t:

$$Y_t = X_t + W_t = u_t + \alpha u_{t-1} + v_t + \beta v_{t-1}$$

We know (by Hamilton, 1994) that this is an MA(1) process.

Now, if we compare it with the MA-terms in the conditional expected value of the cross-sectionally aggregated process (Y_t):

$$E(Y_t | Y_{t-1}, Y_{t-2}) = (\alpha + \beta) Y_{t-1} + (-\alpha\beta) Y_{t-2} + \lambda_u - \beta\lambda_u + \lambda_v - \alpha\lambda_v$$

It can be seen that this is the same as the conditional expected value of sum of two MA(1) processes which is an MA(1) process. Therefore, the sum of the conditional expected values of two INMA(1) processes is equal to the conditional expected value of an INMA(1) process.

Finding the parameters of the aggregated process

Consider the case of adding two independent INMA(1) processes $X_t + W_t$. X_t has the following stochastic characteristics:

$$X_t = u_t + \alpha \circ u_{t-1}$$

$$E(X_t) = \text{var}(X_t) = (1+\alpha)\lambda_u \text{ and } E(X_t X_{t-k}) = \alpha\lambda_u$$

and W_t has stochastic properties as follows:

$$W_t = v_t + \beta \circ v_{t-1}$$

$$E(W_t) = \text{var}(W_t) = (1+\beta)\lambda_v \text{ and } E(W_t W_{t-k}) = \beta\lambda_v$$

Now, if we assume that Y_t ($Y_t = X_t + W_t$) is in fact an INMA(1) process, we have:

$$Y_t = X_t + W_t = Z_t + \phi \circ Z_{t-1}$$

where Z_t has Poisson distribution with mean λ_Z. Therefore, we have:

$$E(Y_t) = E(X_t) + E(W_t) = (1+\alpha)\lambda_u + (1+\beta)\lambda_v$$

The RHS of the above equation should be equal to $(1+\phi)\lambda_Z$.

Now, the autocovariance of Y_t is obtained by:

$$\begin{aligned} E(Y_t Y_{t-k}) &= E[(X_t + W_t)(X_{t-k} + W_{t-k})] = E(X_t X_{t-k}) + E(W_t W_{t-k}) \\ &= \alpha\lambda_u + \beta\lambda_v \end{aligned}$$

which should be equal to $\phi\lambda_Z$. So, we have the following two equations:

$$\begin{cases} (1+\alpha)\lambda_u + (1+\beta)\lambda_v = (1+\phi)\lambda_Z \\ \alpha\lambda_u + \beta\lambda_v = \phi\lambda_Z \end{cases}$$

Solving the above set of equations provides us with a unique value for the parameter of the aggregated series, Y_t, in terms of parameters of the original series, X_t and W_t.

$$\phi = \frac{\alpha\lambda_u + \beta\lambda_v}{\lambda_u + \lambda_v}$$

Therefore, we have obtained the parameters of Y_t process in terms of those of the X_t and W_t processes.

3.3.10.4 Over a Forecast Horizon Aggregation

Consider the case in which demand follows a PoINAR(1) process of Equation 3-9. The demand over lead time is therefore given by:

$$\sum_{i=1}^{l+1} Y_{t+i} = (\alpha \circ Y_t + \alpha^2 \circ Y_t + \cdots + \alpha^{l+1} \circ Y_t)$$
$$+(Z_{t+1} + \alpha \circ Z_{t+1} + \cdots + \alpha^l \circ Z_{t+1})$$
$$+ (Z_{t+2} + \alpha \circ Z_{t+2} + \cdots + \alpha^{l-1} \circ Z_{t+2}) + \cdots + (Z_{t+l} + \alpha \circ Z_{t+l}) + Z_{t+l+1}$$

The conditional mean and variance of lead time demand (LTD) are as follows (see Appendix 3.D for the proof)

$$E\left(\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right) = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\lambda}{1-\alpha}\left((l+1) - \sum_{j=1}^{l+1} \alpha^j\right)$$

Equation 3-60

$$\mathrm{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = Y_t \sum_{j=1}^{l+1} \alpha^j (1-\alpha^j) + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$
$$+\frac{2\lambda}{1-\alpha} \sum_{j=1}^{l} \alpha^{2j-1}\left[(l-j+1) - \frac{\alpha(1-\alpha^{l-j+1})}{1-\alpha}\right]$$

Equation 3-61

Here, we want to know the properties of the lead time demand ($\sum_{i=1}^{l+1} Y_{t+i}$). We start with the simplest case of $l = 1$:

$$Y_\tau^L = \sum\nolimits_{i=1}^{2} Y_{t+i} = Y_{t+1} + Y_{t+2}$$

and

$$Y_{\tau+1}^{L} = \sum_{i=1}^{2} Y_{t+i+1} = Y_{t+2} + Y_{t+3}$$

This means that we have overlap and, as stated before, the result of adding two INAR(1) processes with overlap will be an INAR(1) process. The aggregation of an INARMA(p,q) over a lead time, l, will be discussed in Chapter 6. The results of lead time forecasting of INARMA processes are compared to those of benchmark methods (Croston's method, SBA, and SBJ) for synthetic and empirical data (see Chapters 8 and 9).

3.4 Summary of Literature Review

Having reviewed the literature in earlier sections, this section aims to draw together conclusions and to organise them meaningfully. Table 3-2 provides a summary of the reviewed literature on integer autoregressive moving average models.

Reviewing literature on INARMA models reveals the need for working on stochastic characteristics, diagnosis, estimation of parameters and forecasting for INARMA(p,q) models. This is partly done in this chapter as the unconditional second moment and the autocorrelation function of the process have been found. The partial autocorrelation structure of INARMA(p,q) models has also been found and will be presented in the next chapter.

The aggregation of Guassian ARIMA processes has been discussed in many papers. However, the same is not true for INARMA processes. The work by Brännäs et al. (2002) on temporal aggregation of INAR processes is criticized in this research. Some new results have been found for cross-sectional aggregation. The results of aggregation of an INARMA(p,q) process over a lead time and the conditional first moment of the aggregated process will be presented in chapter 6.

Table 3-2 Literature survey on Integer Autoregressive Moving Average models

	INAR(1)	INAR(2)	INAR(*p*)	INMA(1)	INMA(2)	INMA(*q*)	INARMA(*p*,*q*)
Stochastic Characteristics	McKenzie 1985 McKenzie 1986 Al-Osh & Alzaid 1987 Alzaid and Al-Osh, 1988 McKenzie 1988 Alzaid and Al-Osh, 1993 Brännäs, 1995 Brännäs and Hellström, 2001 Brännäs et al., 2002 Jung and Tremayne, 2003 Silva and Oliveira, 2004 Silva et al., 2005 Jung and Tremayne, 2006b Zheng et al., 2007 Weiß, 2008c	Jung and Tremayne, 2003 Jung and Tremayne, 2006a	Alzaid and Al-Osh, 1990 Du and Li, 1991 Latour, 1998 Silva and Oliveira, 2005 Aly and Bouzar, 2005 Zheng et al., 2006 Kim and Park, 2008 Bu and McCabe, 2008 Bu et al., 2008 Drost et al., 2008 Weiß, 2008a	McKenzie 1988 Al-Osh and Alzaid, 1988 Alzaid and Al-Osh, 1993 Aly and Bouzar, 1994 Brännäs et al., 2002 Jung and Tremayne, 2003	Jung and Tremayne, 2006a	McKenzie 1988 Al-Osh and Alzaid, 1988 Alzaid and Al-Osh, 1993 Aly and Bouzar, 1994 Brännäs and Hall, 2001 Brännäs and Quoreshi, 2004 Weiß, 2008b	McKenzie 1988 PoINARMA(1,*q*) Alzaid and Al-Osh, 1993 (INARMA(*p*,*p*-1) with quasi binomial thinning) Aly and Bouzar, 1994 (INARMA(1,*q*) with a general thinning) Neal and Rao, 2007 (likelihood of INARMA(*p*,*q*) via MCMC)

Table 3-2 *contd.*

	INAR(1)	INAR(2)	INAR(p)	INMA(1)	INMA(2)	INMA(q)	INARMA(p,q)
Identification	Testing for independence: Jung and Tremayne, 2003; 2006b						
	Freeland and McCabe, 2004a		Bu and McCabe, 2008			Brännäs and Quoreshi, 2004	Enciso-Mora et al., 2008 (Reverse Jump MCMC algorithm for order selection)
Estimation	Al-Osh & Alzaid 1987 (YW, CLS, CML) Franke and Seligmann, 1993 (CML) Franke and Subba Rao, 1995 (CML) Brännäs, 1994 (CLS, ML, GMM) Brännäs, 1995 (CLS, weighted CLS, and GMM) Park and Oh, 1997 (YW and CLS) Brännäs and Hellström, 2001 (CLS, Weighted CLS, GMM) Brännäs et al., 2002 (CLS) Silva and Oliveira, 2004 (Whittle criterion and CML)	Jung and Tremayne, 2006a (YW) Jung and Tremayne, 2006b (YW)	Du and Li, 1991 (YW, CLS) Latour, 1998 (YW and CLS) Silva and Oliveira, 2005 (Whittle criterion) Zheng et al., 2006 (ML, CLS, MQL and GMM for RCINAR(p)) Kim and Park, 2008 (YW, CLS) Bu et al., 2008 (CLS, CML) Drost et al., 2008 (asymptotically efficient estimator for parametric INAR(p) models)	Jung and Tremayne, 2006a (YW)	Jung and Tremayne, 2006a (YW)	Brännäs and Hall, 2001 (YW, CLS, GMM based on pgf) Brännäs and Quoreshi, 2004 (CLS, FGLS, GMM)	Neal and Rao, 2007 (ML via MCMC)

Table 3-2 *contd.*

	INAR(1)	INAR(2)	INAR(*p*)	INMA(1)	INMA(2)	INMA(*q*)	INARMA(*p*,*q*)
Estimation	Silva et al., 2005 Replicated INAR(1) processes (YW, CLS, IWCLS, CML, Whittle, Bayesian) Silva & Oliveira 2005 (Whittle criterion) Freeland and McCabe, 2005 (CLS) Jung and Tremayne, 2006b (YW, CLS, ML) Zheng et al., 2007 (ML, CLS and modified quasi-likelihood (MQL) for RCINAR(1) model)						
Forecasting	Freeland and McCabe, 2004b (Coherent forecasting using the median of the *h*-step ahead conditional distribution) McCabe & Martin 2005 (Coherent forecasts with Bayesian methodology)	Jung and Tremayne, 2006b (Coherent forecasting using bootstrap)	Du and Li, 1991 (MMSE forecasts) Kim and Park, 2008 (Coherent forecasting using bootstrap) Bu and McCabe, 2008 (Coherent forecasting using a Markov Chain approach)	Brännäs et al., 2002 (MMSE forecasts)		Brännäs and Hall, 2001 (MMSE forecasts) Brännäs and Quoreshi, 2004 (MMSE forecasts)	Neal and Rao, 2007 (*h*-step ahead prediction distribution of INARMA(*p*,*q*) via MCMC)

3.5 Conclusions

The literature on intermittent demand forecasting supports the need for a model-based method. Time series models for count data provide such methods but not all these models have practical properties for IDF. As an example of such models, discrete autoregressive moving average (DARMA) models have been reviewed. These models have their main application in the hydrological literature because a realization of a DARMA process contains many runs of a constant value.

Another class of count data time series models are called integer autoregressive moving average (INARMA) models. These models do not have the problem of DARMA models and have interesting properties such as having the same correlation structure as ARMA models.

Different INARMA models have been introduced and some of their statistical properties have been reviewed. For higher order integer autoregressive and moving average models there are two approaches regarding the thinning mechanisms. In this study we follow the approach that assumes independence of thinning operations of each other and all the previous corresponding operations.

It has been found that the statistical properties of the general INARMA(p,q) model (e.g. the autocorrelation function) have not been looked at in the literature. The unconditional variance and the ACF of an INARMA(p,q) process have been found in this chapter.

INARMA models have had applications in different areas such as medical science and economics. Although all of these applications focus on the direct physical interpretation of these processes as a birth and death process, these models should not only be restricted to such situations. Our study suggests a new practical area for INARMA processes in modelling and forecasting intermittent demand.

Finally, different schemes of aggregation of an INAR(1) process have been discussed, namely temporal, cross-sectional, and lead time aggregation. The results of aggregation of an INAR(1) process over lead time have been presented. Similar results for an INARMA(p,q) process will be presented in chapter 6.

Chapter 4 IDENTIFICATION IN INARMA MODELS

4.1 Introduction

This chapter reviews methods for identification of the autoregressive and moving average order of an INARMA model. Different tests suggested for investigating serial dependence in time series of counts are reviewed in section 4.2. The autocorrelation function (ACF) and partial autocorrelation function (PACF) can be used to select the appropriate INARMA models, similar to ARMA models. The ACF and PACF structures of INARMA models are reviewed in section 4.3. The residual analysis to check the model adequacy is discussed in section 4.4. The Akaike information criterion (AIC) is an easily automated method of identification and is introduced in section 4.5. The identification procedures undertaken in this research are discussed in section 4.6. Finally, the conclusions are summarized in section 4.7.

4.2 Testing Serial Dependence

Jung and Tremayne (2003) argue that when analyzing any time series of counts, the first natural step is to investigate if the data exhibit significant serial dependence. This is because if the data do not show such dependence, there is no need for INARMA methods. In such cases, methods for independent data should be used.

In this section, methods of testing for independence in a time series of counts are presented. Several methods have been suggested in the literature, which we briefly review.

4.2.1 Runs Test

In this test, the original series is dichotomized on the basis of some criterion. The median is often recommended for this test, and the observations that are identical to the sample median are discarded. However, in stationary time series of low value counts one can expect to see the median value very frequently; therefore many observations would have to be eliminated and this affects the power of the test. Instead, Jung and Tremayne (2003) use the sample mean as they argue that it will rarely be integer valued.

Under the null hypothesis (H_0) of no serial dependence, the distribution of the number of runs of a time series $(Y_1, Y_2, \ldots, Y_T)$ is tested by the following statistic (Jung and Tremayne, 2006a):

$$Z = \frac{R - 1 - [2T_1(T - T_1)]/T}{\left\{\frac{2T_1(T - T_1)[2T_1(T - T_1) - T]}{[T^2(T - 1)]}\right\}^{1/2}}$$

Equation 4-1

where R is the number of runs and T_1 is the number of positive runs. Wald and Wolfowitz (1940) show that $Z \overset{d}{\rightarrow} N(0,1)$ under the null hypothesis, where $\overset{d}{\rightarrow}$

indicates convergence in ditrubution. Therefore, the null hypothesis is rejected if $Z < z_\alpha$, where z_α is the relevant quantile of the standard normal distribution and α is the type I error. The test is one-sided because INARMA models only have positive autocorrelations.

4.2.2 The Score Test

The score test of Freeland (1998) is another approach of testing for independence in a time series of counts. Under the null hypothesis (H_0) of the underlying time series Y_t being i.i.d. Poisson variables, Freeland (1998) shows that the test statistic is:

$$S = \frac{1}{\bar{Y}\sqrt{n}} \sum_{t=2}^{n} (Y_{t-1} - \bar{Y})(Y_t - \bar{Y})$$

Equation 4-2

where $\bar{Y}$ is the sample mean $(\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t)$ and n is the number of observations. The above statistic can be approximated by the standard normal distribution, $S \xrightarrow{d} N(0,1)$ and the null hypothesis is rejected if $S > z_\alpha$.

Jung and Tremayne (2003) provide a modified statistic based on the mean-variance equality property of the Poisson distribution. The modified statistic is given by:

$$S^* = \sqrt{n} \frac{\sum_{t=2}^{n} (Y_{t-1} - \bar{Y})(Y_t - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

Equation 4-3

which is asymptotically equivalent to the S-statistic under the null hypothesis. Here, again the null hypothesis is rejected if $S^* > z_\alpha$.

4.2.3 Portmanteau-type Tests

Jung and Tremayne (2003) use two portmanteau-type tests originally designed by Venkataraman (1982) and Mills and Seneta (1989) to measure the goodness-of-fit in branching processes as tests for independence in a time series of counts.

Under the null hypothesis (H_0) of no serial dependence or i.i.d. random variables, the modified version of the statistic presented by Venkataraman (1982) is:

$$Q_{acf}(k) = \frac{\sum_{i=1}^{k} \hat{Q}_{i+1}^2 \left[\sum_{t=1}^{n}(Y_t - \bar{Y})^2\right]^2}{\sum_{t=i+2}^{n}(Y_t - \bar{Y})^2 (Y_{t-i-1} - \bar{Y})^2}$$

Equation 4-4

where $\hat{Q}_k = \sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y}) / \sum_{t=1}^{n}(Y_t - \bar{Y})^2$ and $k \geq 1$ is an arbitrary integer (in their Monte-Carlo study, Jung and Tremayne (2003) use $k = 1, 5, 10$). Under the null hypothesis of i.i.d. Poisson variables $\{Y_t\}$, $Q_{acf}(k) \xrightarrow{d} \chi^2(k)$ as $n \to \infty$.

The second test is an adapted version of the test presented by Mills and Seneta (1989) with the statistic:

$$Q_{pacf}(k) = \frac{\sum_{i=1}^{k} \hat{\Phi}_{i+1}^2 \left[\sum_{t=1}^{n}(Y_t - \bar{Y})^2\right]^2}{\sum_{t=i+2}^{n}(Y_t - \bar{Y})^2 (Y_{t-i-1} - \bar{Y})^2}$$

Equation 4-5

It can be easily seen that this is exactly the same as $Q_{acf}(k)$ except that $\hat{Q}_{i+1}$ is replaced by $\hat{\Phi}_{i+1}$. Here, $\hat{\Phi}_k$ is the kth order sample partial autocorrelation and k is as defined before. Similarly, $Q_{pacf}(k) \xrightarrow{d} \chi^2(k)$ as $n \to \infty$.

It should be noted that the first-order lag sample correlations are ignored in both statistics. These two tests can be used to distinguish between INAR(1) and INMA(1) structures.

Two portmanteau tests used in the ARMA literature to find if the data has any serial dependence are the Box-Pierce and the Ljung-Box tests. The latter is an enhancement of the former to improve the performance of the test for small sample sizes (Ljung and Box, 1978). The Ljung-Box statistic is given by:

$$Q^* = n(n+2) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{n-j}$$

Equation 4-6

where n is the number of observations and $\hat{\rho}_j$ is the sample autocorrelation at lag j. A large value of Q^* indicates that the model is inadequate.

4.3 Identification based on ACF and PACF

The sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) have been widely used in the literature for identification of the autoregressive and moving average order of the INARMA models (Latour, 1998; Jung and Tremayne, 2006a; Zheng et al., 2006; Zhu and Joe, 2006; Bu and McCabe, 2008). In this section, we review these functions for different INARMA models.

4.3.1 Autocorrelation Function (ACF)

The Autocorrelation Function (ACF) is defined as a plot of the autocorrelations at lag k versus the lag k. In this section, we investigate the autocorrelation function of an INARMA(p,q) process. First, we recall that the ACF of an INAR(p) process of:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \cdots + \alpha_p \circ Y_{t-p} + Z_t$$

is determined by Du and Li (1991) as:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p}$$

Equation 4-7

It can be seen that the correlation structures of INAR(p) and AR(p) processes are the same. For an INMA(q) process of:

$$Y_t = Z_t + \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2} + \cdots + \beta_q \circ Z_{t-q}$$

it is shown by Brännäs and Hall (2001) that the ACF for an INMA(q) process with Poisson marginal distribution is given by:

$$\rho_k = \begin{cases} \frac{\sum_{j=0}^{q-k} \beta_j \beta_{j+k}}{\sum_{j=0}^{q} \beta_j} & \text{for } k = 1, \dots, q \\ 0 & \text{for } k > q \end{cases}$$

Equation 4-8

Again, it can be seen that this is analogous to the ACF of an MA(q) process.

In chapter 3, we showed that the autocorrelation function of an INARMA(p,q) process is as follows:

$$\rho_k = \begin{cases} \dfrac{\alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} + \beta_k\gamma_0^{YZ} + \beta_{k+1}\gamma_1^{YZ} + \cdots + \beta_q\gamma_{q-k}^{YZ}}{\gamma_0} & k \le q \\ \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p} & k > q \end{cases}$$

Equation 4-9

When $k \geq q+1$, the autocorrelation is:

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p} \qquad \text{for } k \geq q+1$$

Equation 4-10

Therefore, the ACF of an INARMA(p,q) process is analogous to that of an ARMA(p,q) process and it can be used in identifying the integer-valued time series.

4.3.2 Partial Correlation Function (PACF)

The correlation between two variables can be used as a measure of interdependence. However, when a variable is correlated with a second variable, this may be due to the fact that they both are correlated with another variable(s). Therefore, it may be of interest to examine the correlations between variables when other variables are held constant. These are called *partial correlations* (Hamilton, 1994).

PACF is a device to identify the autoregressive order of a stationary time series. It has been shown that, although an AR(p) process has an infinite ACF, it can be described in terms of p non-zero functions of the autocorrelations (Box et al., 1994).

In this section, we examine the partial autocorrelation function of INARMA processes. The section is organized as follows. First the PACF of an INAR(p) process is studied. The PACF of INMA(q) and INARMA(p,q) processes are then discussed.

4.3.2.1 PACF of an INAR(p) Model

In this section, we examine the partial autocorrelation function of INAR processes. The INAR(p) process is defined by the recursion:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \cdots + \alpha_p \circ Y_{t-p} + Z_t$$

Here we assume that Z_t has a Poisson distribution with parameter λ. Multiplying the above equation by Y_{t-k} produces:

$$Y_t Y_{t-k} = Y_{t-k}(\alpha_1 \circ Y_{t-1}) + Y_{t-k}(\alpha_2 \circ Y_{t-2}) + \cdots + Y_{t-k}(\alpha_p \circ Y_{t-p}) + Y_{t-k} Z_t$$

Equation 4-11

but we know that:

$$E(Y_t) = \alpha_1 E(Y_{t-1}) + \alpha_2 E(Y_{t-2}) + \cdots + \alpha_p E(Y_{t-p}) + E(Z_t)$$

$$\mu = \mu(\alpha_1 + \alpha_2 + \cdots + \alpha_p) + \lambda$$

So, if we take the expected value of Equation 4-11 and subtract $E(Y_t)E(Y_{t-k})$ from it, we have:

$$\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_p \gamma_{k-p} + \mathrm{cov}(Y_{t-k}, Z_t) \qquad k \geq 0$$

Equation 4-12

considering the fact that $E\left[Y_{t-k}\left(\alpha_j \circ Y_{t-j}\right)\right] = \alpha_j E(Y_{t-k} Y_{t-j})$ (Silva and Oliveira, 2004). The last term in the RHS is zero because Y_{t-k} can only involve innovation terms up to time $t-k$ and therefore is uncorrelated with Z_t. Dividing the Equation 4-12 by γ_0 yields:

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p} \qquad k \geq 0$$

Equation 4-13

which is analogous to the difference equation for Gaussian AR processes. The *Yule-Walker* equations are:

$$\begin{array}{lllllllll}
\rho_1 & = & \alpha_1 & + & \alpha_2 \rho_1 & + & \cdots & + & \alpha_p \rho_{p-1} \\
\rho_2 & = & \alpha_1 \rho_1 & + & \alpha_2 & + & \cdots & + & \alpha_p \rho_{p-2} \\
\vdots & & \vdots & & \vdots & & \cdots & & \vdots \\
\rho_p & = & \alpha_1 \rho_{p-1} & + & \alpha_2 \rho_{p-2} & + & \cdots & + & \alpha_p
\end{array}$$

Equation 4-14

The autocorrelation function of an INAR(p) process can be described in terms of p

nonzero functions of the autocorrelations.

In an integer autoregressive process of order k:

$$\rho_j = \alpha_{k1}\rho_{j-1} + \alpha_{k2}\rho_{j-2} + \cdots + \alpha_{k(k-1)}\rho_{j-k+1} + \alpha_{kk}\rho_{j-k} \qquad \text{for } j = 1,2,\dots,k$$

Equation 4-15

where α_{kj} is the jth coefficient in an integer autoregressive representation of order k. Therefore α_{kk} is the last coefficient and $\alpha_{kj} = 0$ for $j > k$.

The Equation 4-15 leads to the Yule-Walker equations which may be written as:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_{k1} \\ \alpha_{k2} \\ \vdots \\ \alpha_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

Equation 4-16

As defined by Hamilton (1994), the kth population partial autocorrelation (denoted by α_{kk}) is the last coefficient in a linear (for Gaussian ARMA processes) projection of Y on its k most recent values. In the case of integer autoregressive models, it can be stated as:

$$Y_t = \alpha_{k1} \circ Y_{t-1} + \alpha_{k2} \circ Y_{t-2} + \cdots + \alpha_{kk} \circ Y_{t-k}$$

Equation 4-17

which results in the same sets of difference equations based on autocorrelations. The justification for using partial autocorrelations is that, if data really were generated by an INAR(p) process, only the p most recent values of Y would be useful for forecasting and the coefficients on Y's more than p periods in the past are equal to zero, which means that:

$$\alpha_{kk} = 0 \qquad \text{for } k = p + 1, p + 2, \dots$$

We can express the INAR(p) process as follows:

$$Z_t = Y_t - \alpha_1 \circ Y_{t-1} - \alpha_2 \circ Y_{t-2} - \cdots - \alpha_p \circ Y_{t-p} = Y_t - \sum_{j=1}^{p} \alpha_j \circ Y_{t-j}$$

It can be seen that the series in the RHS of the above equation is finite. Therefore, the

PACF of an INAR (p) process is finite.

Although an INAR(p) process differs from the AR(p) process due to the thinning operations, these two processes share some properties including the ACF and PACF structure. Therefore, using the sample partial autocorrelation function (SPACF) may help us in identifying the autoregressive order of an INAR time series.

4.3.2.2 PACF of an INMA(q) Model

This section focuses on finding the partial autocorrelation function of an INMA process. The INMA(q) process has the following form:

$$Y_t = Z_t + \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2} + \cdots + \beta_q \circ Z_{t-q}$$

We again assume that Z_t has a Poisson distribution with parameter λ. Using the same argument as for an INAR(p) process, it can be seen that the autocovariance function of an INMA(q) process is:

$$\gamma_k = E[(Z_t + \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2} + \cdots + \beta_q \circ Z_{t-q}).$$
$$(Z_{t-k} + \beta_1 \circ Z_{t-k-1} + \beta_2 \circ Z_{t-k-2} + \cdots + \beta_q \circ Z_{t-k-q})]$$

so

$$\gamma_k = \begin{cases} (\beta_k + \beta_1\beta_{k+1} + \beta_2\beta_{k+2} + \cdots + \beta_{q-k}\beta_q)\sigma_Z^2 & k = 1,2,\dots,q \\ 0 & k > q \end{cases}$$

Equation 4-18

Thus, the autocorrelation function (ACF) is given by:

$$\rho_k = \begin{cases} \dfrac{\beta_k + \beta_1\beta_{k+1} + \beta_2\beta_{k+2} + \cdots + \beta_{q-k}\beta_q}{1 + \beta_1 + \cdots + \beta_q} & k = 1,2,\dots,q \\ 0 & k > q \end{cases}$$

Equation 4-19

It can be seen that the autocorrelation function of an INMA(q) process is zero, beyond the order q of the process.

As discussed earlier in section 4.3.2.1, if the data were generated by an INAR(p)

process, only the p most recent values of Y would be useful for forecasting and therefore the PACF cuts off after lag p. However, if the data were generated by an INMA(q) process with $q \geq 1$, then the partial autocorrelation α_{kk} asymptotically approaches zero instead of cutting off abruptly.

The PACF of the process (α_{kk}) can be obtained by solving the set of equations given by Equation 4-15 using the ACF of the process. For example, Al-Osh and Alzaid (1988) find the first and second partial autocorrelations of an INMA(1) process as:

$$\alpha_{11} = \rho_1$$

$$\alpha_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{-\beta^2}{1 + 2\beta}$$

which can be seen is the same as that of a MA(1) process.

4.3.2.3 PACF of an INARMA(*p*,*q*) Model

The partial autocorrelation function of an INARMA(p,q) process is examined in this section. This process satisfies the difference equation:

$$Y_t = \sum_{j=1}^{p} \alpha_j \circ Y_{t-j} + Z_t + \sum_{j=1}^{q} \beta_j \circ Z_{t-j}$$

An INARMA(p,q) process can be written in form of (see Appendix 4.A):

$$Z_t = Y_t - \sum_{j=1}^{\infty} \sum_{i=1}^{n_j} \psi_{ij} \circ Y_{t-j}$$

Equation 4-20

where $n_j = \begin{cases} \left(\sum_{i=1}^{q} n_{j-i}\right) + 1 & 0 < j \leq p \\ \sum_{i=1}^{q} n_{j-i} & j > p \end{cases}$.

The values of ψ_{ij} can be found by repeated multiplications of the parameters of INARMA(p,q) process.

It can be seen that the series in the RHS of the Equation 4-20 is infinite. Therefore, the PACF of an INARMA(p,q) process, similar to an INMA(q) process, is infinite.

As shown in section 4.3.1, the structure of the autocorrelation function of the INARMA(p,q) process is analogous to that of an ARMA process. The structure of the partial autocorrelation function of this process is also similar to that of an ARMA process, i.e. the PACF of a mixed integer autoregressive moving average process is infinite and it has the same shape as the PACF of a pure integer moving average process.

4.4 Residual Analysis

The residuals of the INARMA models can provide a check of model adequacy (Jung and Tremayne, 2006b). After identification of the appropriate model using ACF and PACF and estimation of the parameters of the identified model, the residuals should be examined to check for any serial dependence. Any dependence in the residuals would suggest that a different model should be used. The ACF and PACF of the residuals should be plotted for this reason.

Freeland and McCabe (2004a) define two sets of residuals: one for the arrivals component and another for the continuation process of a PoINAR(p) process. The residuals for the continuation component are given by:

$$r_{1t} = \alpha_j \circ Y_{t-j} - \alpha_j Y_{t-j} \qquad \text{for } t = p+1, \dots, n$$

Equation 4-21

The residuals for the arrivals component are:

$$r_{2t} = Z_t - \lambda$$

Equation 4-22

However, as they mention, these definitions are not practical because $\alpha_k \circ Y_{t-k}$ and Z_t cannot be observed and should be replaced with their conditional expected values. Adding the new components then results in the usual definition of residuals for a PoINAR(p) process:

$$r_t = Y_t - \sum_{i=1}^{p} \alpha_i Y_{t-i} - \lambda$$

Equation 4-23

Bu and McCabe (2008) suggest that in order to check the adequacy of the selected model, checking the traditional residual of the Equation 4-23 is not enough and the components residuals (r_{1t} and r_{2t}) should also be tested to examine the suitability of each component in the model. The expected value of $\alpha_k \circ Y_{t-k}$ and Z_t are provided in terms of the conditional probabilities. This resolves the impracticality issue mentioned in Freeland and McCabe (2004a).

The residuals of an INMA(q) process are given by (Brännäs and Hall, 2001):

$$r_t = Y_t - \lambda - \sum_{i=1}^{q} \beta_i Z_{t-i}$$

Equation 4-24

Based on Equation 4-23 and Equation 4-24, the residuals of an INARMA(p,q) model can be obtained from:

$$r_t = Y_t - \sum_{i=1}^{p} \alpha_i Y_{t-i} - \lambda - \sum_{i=1}^{q} \beta_i Z_{t-i}$$

Equation 4-25

It will be explained in section 4.5 that the identification methods used in this PhD thesis are not based on ACF and PACF. However, as an example, here we show what the ACF and PACF of one INAR(1) series look like. The series is selected from those of the 16,000 series data set of chapter 9. The time series plot for all 72 periods is provided in Figure 4-1. The sample ACF and sample PACF of the above series are presented in Figure 4-2. The sample PACFs suggest that an INAR(1) model is appropriate.

The parameters of the identified INAR(1) model are then estimated using the Yule-Walker (YW) estimation method (see section 5.3.1) to be $\hat{\alpha} = 0.5581$ and $\hat{\lambda} = 0.0552$.

The next step is to check the model's adequacy using the residual analysis. The residuals of the INAR(1) model are defined as:

$$\hat{\varepsilon}_t = Y_t - \hat{\alpha}_{YW} Y_{t-1} - \hat{\lambda}_{YW}$$

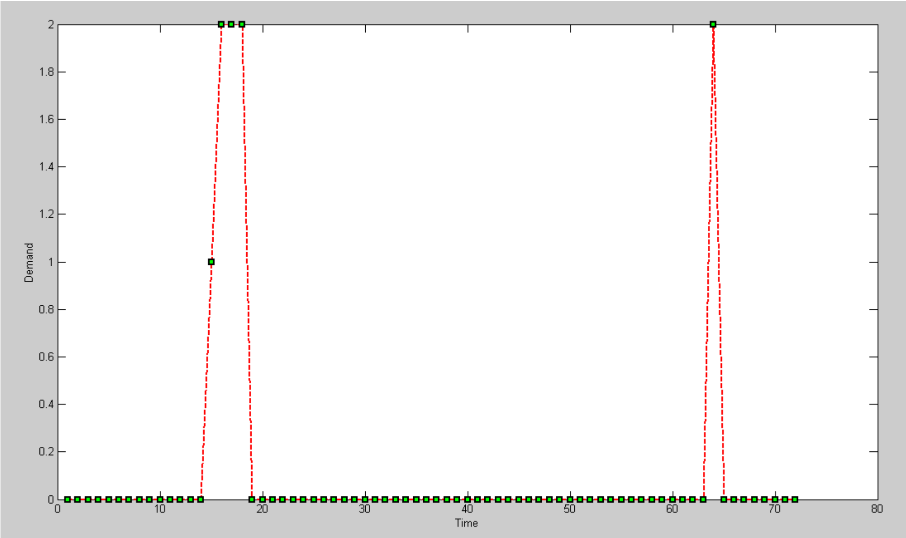

Figure 4-1 Time series plot of one demand series among 16,000 series

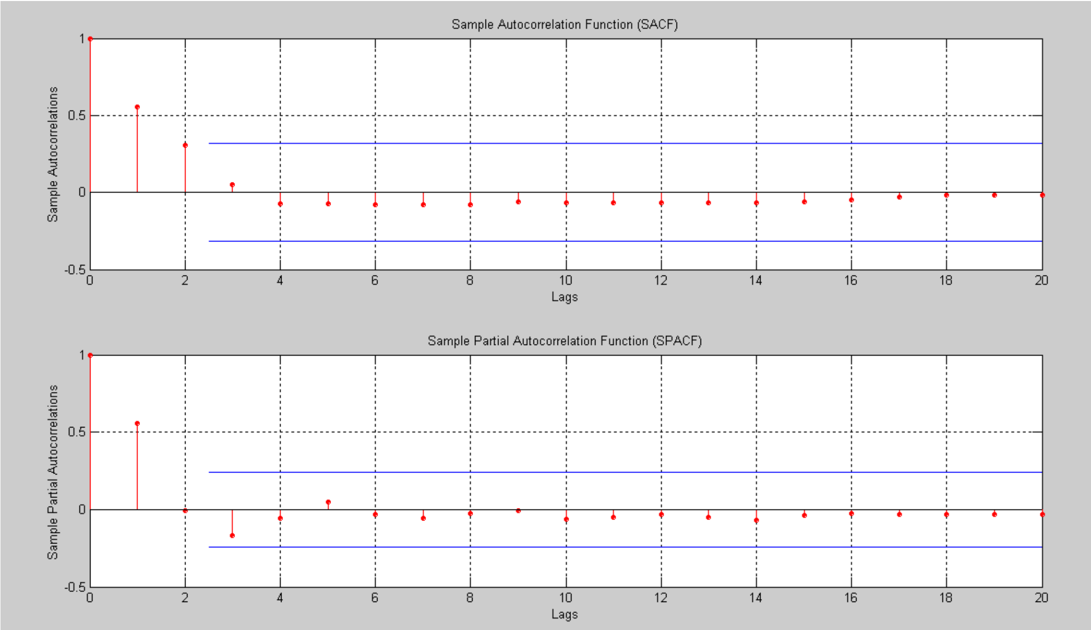

Figure 4-2 Correlograms of the selected series among 16,000 series

where $\hat{\alpha}_{YW}$ and $\hat{\lambda}_{YW}$ are the Yule-Walker estimates of parameters in the INAR(1) model.

If any dependence structure exists in the residuals, a different model specification example would be considered. In order to check if such dependence exists in our example, the SACFs and SPACFs of the residuals of the estimated INAR(1) model are depicted in Figure 4-3. The figure suggests that there is no obvious dependence structure left in the residuals.

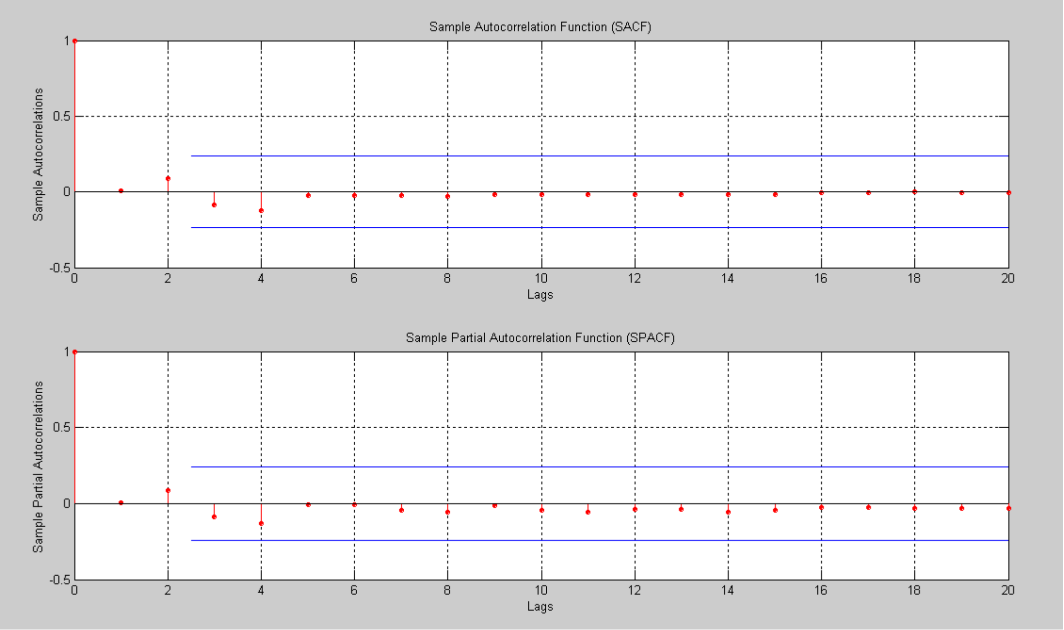

Figure 4-3 Correlograms of the residuals of the INAR(1) model

4.5 Identification based on Penalty Functions

It has been argued that distinguishing between autoregressive moving average models based on the Box-Jenkins procedure is difficult (Chatfield and Prothero, 1973; Newbold and Granger, 1974). This is because the ACF and PACF plots cannot easily identify mixed ARMA models. Moreover, the identification of ARMA models usually involves subjective judgment. The same is true for identification of INARMA models using ACF and PACF.

The Kullback-Leibler information (Kullback and Leibler, 1951) is used to measure the difference between two probability density functions $f(x)$ and $g(x)$:

$$I(f,g) = \int f(x)\log\left(\frac{f(x)}{g(x|\theta)}\right)dx$$

Equation 4-26

In the above equation, $I(f,g)$ denotes the information lost when g is used to approximate f. f is considered to be fixed and g varies over θ.

Akaike (1973) introduces the Akaike information criterion (AIC) as an approximately unbiased estimate of Kullback-Leibler information. The AIC is given by:

$$\mathrm{AIC} = (-2)\log(\text{maximum likelihood}) + 2m$$

Equation 4-27

where m is the number of estimable parameters in the approximating model. Ozaki (1977) shows that the AIC of the ARMA(p,q) model is given by:

$$\mathrm{AIC} \approx N\log\hat{\sigma}_a^2 + 2m$$

Equation 4-28

where $\hat{\sigma}_a^2$ is the residual variance and $m = p + q + 1$. When the sample size is small $(N/m < 40)$, the above expression is biased and the following bias correction is necessary (Hurvich and Tsai, 1989; Hurvich and Tsai, 1995):

$$\mathrm{AIC_C} \approx N\log\sigma_a^2 + 2m + 2m(m+1)/(N-m-1)$$

Equation 4-29

AIC has also been used in the INARMA literature (see for example: Böckenholt, 1999; Brandt et al., 2000; Zhu and Joe, 2006). However, the complexity of the likelihood function of these models has led to some limitations in the use of AIC.

The likelihood function of an INAR(p) process with Poisson innovations is given by (Bu, 2006):

$$L(\alpha_1, \dots, \alpha_p, \lambda) = \prod_{t=p+1}^{n} P(Y_t | Y_{t-1}, \dots, Y_{t-p})$$

Equation 4-30

where the conditional probability function $P(Y_t | Y_{t-1}, \dots, Y_{t-p})$ is:

$$P(Y_t | Y_{t-1}, \dots, Y_{t-p}) = \sum_{i_1=0}^{\min(Y_{t-1}, Y_t)} \binom{Y_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{Y_{t-1}-i_1}$$

$$\sum_{i_2=0}^{\min(Y_{t-2}, Y_t - i_1)} \binom{Y_{t-2}}{i_2} \alpha_2^{i_1} (1-\alpha_2)^{Y_{t-2}-i_2} \dots$$

$$\sum_{i_p=0}^{\min[Y_{t-p}, Y_t - (i_1 + \cdots + i_{p-1})]} \binom{Y_{t-p}}{i_p} \alpha_p^{i_p} (1-\alpha_p)^{Y_{t-p}-i_p} \frac{e^{-\lambda} \lambda^{Y_t - (i_1 + \cdots + i_{p-1})}}{[Y_t - (i_1 + \cdots + i_{p-1})]!}$$

Equation 4-31

It can be seen that the logarithm of Equation 4-31 cannot be simplified as in ARMA models. It should also be mentioned that the likelihood function of the INARMA(p,q) process is not established yet.

4.6 The Identification Procedure

For the simulation experiment, we need an automated method for identification that can be used for thousands of replications. Because the likelihood function of an INARMA(p,q) process is not yet found and even the likelihood function of an INAR(p) process is very complicated, the penalty functions for these models are not easy to find. It has been shown in chapter 3 that an INAR process is analogous to an AR process in autocorrelation structure and also forecasting. Therefore, it has been suggested that the standard programmes for AR processes, which are mainly based on AIC or BIC, could also be used for INAR processes (Latour, 1998). The same is true for an INMA process regarding the ACF structure and the conditional expected value.

As mentioned in the previous section, the use of AIC in the INARMA literature is limited to those processes for which likelihood functions have been derived. This excludes the INMA and mixed models. Based on the above argument by Latour (1998) we test the performance of AIC of Equation 4-28 (or where applicable, AIC_C of Equation 4-29) for INARMA models.

As discussed in section 4.2, Jung and Tremayne (2003) suggest that in analysing the time series of counts, any serial dependence should first be detected. If no such dependence is found, the complicated INARMA methods can be replaced with easier methods for independent data. Based on the above argument, we use two identification procedures in this thesis: a two-stage and a one-stage method.

In the two-stage method, the first stage distinguishes between the INARMA(0,0) and the other INARMA models. The Ljung-Box statistic of Equation 4-6 is used for this reason. This is because it is a standard test used for conventional ARMA models. Therefore, it is included in most software packages (including MATLAB which is used in this thesis) and, based on the argument by Latour (1998), it can be used for

INARMA models as well. It will be shown in chapter 8 that the rejection percentages under the null hypothesis of i.i.d. Poisson are comparable to the results of the tests suggested by Jung and Tremayne (2003).

The AIC of Equation 4-28 is then used for identification among the other INARMA models. This is again based on the argument of Latour (1998) to use the standard programmes for ARMA models for INARMA models. It should also be mentioned that the AIC of ARMA models has been used in the INARMA literature (e.g. Brännäs and Quoreshi, 2004). The reliability of this identification procedure will be tested in the simulation chapter. To our knowledge, this has not been done in the literature before. The impact of misidentification on the forecast accuracy will also be checked. The one-stage method only uses the AIC of Equation 4-28 (or AIC_C of Equation 4-29) to identify the most appropriate model.

The performance of these two methods will be compared. This will be done in terms of the percentage of time that the correct model is identified and also the accuracy of forecasts based on each method.

4.7 Conclusions

In this chapter, the methods of identification of the autoregressive and moving average orders of an INARMA process have been reviewed. It has been shown that the autocorrelation and partial autocorrelation functions of an INAR(p) process have the same structure as those of an AR(p) process. The same is true for INMA(q) and INARMA(p,q) processes. Therefore, the estimates of the functions (SACF and SPACF) can identify the moving average and autoregressive orders, respectively. The residuals of the estimated INARMA process then need to be checked for any remaining correlations.

Two identification procedures will be used in this thesis. A two-stage method is based on first using the Ljung-Box statistic to identify any correlation in the data series. The next step involves using the AIC of ARMA models to select from the other possible INARMA models. The performance of this procedure in terms of the percentage of time that the model is identified correctly and also the effect of misidentification on forecast accuracy will be tested in chapter 8. To our knowledge, this has not been done in INARMA literature. The two stage method will then be compared to a one-stage identification method based on using the AIC to select among the INARMA models including INARMA(0,0).

Chapter 5 ESTIMATION IN INARMA MODELS

5.1 Introduction

Having identified the appropriate INARMA model, we then need to estimate the parameters of the selected model. Different estimation methods have been used in the literature to estimate the parameters of INAR(p) and INMA(q) models, namely:

- Yule-Walker (YW)
- Conditional least squares (CLS)
- Maximum likelihood (ML)
- Generalized method of moments (GMM)

Each of these methods is briefly reviewed. Table 5-1 lists the main studies on estimation of parameters of INARMA models.

Table 5-1 Research papers on estimation of parameters of INARMA models

Model	YW	CLS	ML	GMM
INAR(1)	Al-Osh and Alzaid, 1987	Al-Osh and Alzaid, 1987	Al-Osh and Alzaid, 1987	Brännäs, 1994 Brännäs and Hellström, 2001
INAR(*p*)	Du and Li, 1991 Jung and Tremayne, 2006b for INAR(2)	Du and Li, 1991	Bu et al., 2008	
INMA(1)	Brännäs and Hall, 2001	Brännäs and Hall, 2001		Brännäs and Hall, 2001
INMA(*q*)	Brännäs and Hall, 2001 for INMA(2)	Brännäs and Hall, 2001		Brännäs and Hall, 2001

The YW method is based on using the Yule-Walker equations of:

$$\begin{array}{lllllllll} \rho_1 & = & \alpha_1 & + & \alpha_2\rho_1 & + & \cdots & + & \alpha_p\rho_{p-1} \\ \rho_2 & = & \alpha_1\rho_1 & + & \alpha_2 & + & \cdots & + & \alpha_p\rho_{p-2} \\ \vdots & & \vdots & & \vdots & & \cdots & & \vdots \\ \rho_p & = & \alpha_1\rho_{p-1} & + & \alpha_2\rho_{p-2} & + & \cdots & + & \alpha_p \end{array}$$

and replacing the theoretical autocorrelations ρ_k by the sample autocorrelations, r_k (Box et al., 1994):

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

Equation 5-1

Lawrence et al. (1978) develop the CLS estimation procedure for stochastic processes based on the minimization of a sum of squared deviations about conditional expectation.

Maximum likelihood estimation method, as the name suggests, finds the parameters that maximize the likelihood of the sample data. The likelihood of the sample data is the probability of obtaining that particular set of data, given the specific probability distribution.

In the generalized method of moments, a set of population moment conditions is first derived based on the assumptions of the model. Then the GMM estimates of the parameters are obtained such that these moment conditions are satisfied for the sample data.

This chapter is organized as follows. The estimate of the parameter of a Poisson INARMA(0,0) process is provided in section 5.2. The YW, CLS, CML, and GMM estimates for the parameters of INAR(1) and INAR(p) processes are reviewed in sections 5.3 and 5.4 (GMM only for INAR(1) process). The corresponding estimation methods for INMA(1) and INMA(q) processes are reviewed in sections 5.5 and 5.6. The YW and CLS estimates of the parameters of an INARMA(1,1) process are derived in section 5.7. Finding the ACF of an INARMA(p,q) model in chapter 3 enables us to find the YW estimates of these models. As an example, the YW estimates of an INARMA(2,2) model are derived in section 5.8. The conclusions are given in section 5.9.

As will be discussed in chapter 7, four INARMA models are selected for simulation and empirical analysis to compete against the benchmark methods. These models are: INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1). The estimation methods used in this thesis for these models are CLS and YW for the last three and also CML for INAR(1). Therefore, these estimates are specifically given in this section. It will be discussed in section 5.2 that the CLS and ML estimation methods result in the same estimate for a Poisson INARMA(0,0) process.

5.2 Estimation in an INARMA(0,0) Model

The INARMA(0,0) process with Poisson marginal distribution is simply an i.i.d. Poisson process of:

$$Y_t = Z_t$$

Equation 5-2

where Z_t are i.i.d. Poisson random variables. The conditional expected value of Y_t given Y_{t-1} is therefore given by:

$$E(Y_t|Y_{t-1}) = \lambda$$

Equation 5-3

where λ is the only parameter to be estimated. The conditional least squares estimate of λ is obtained by minimizing the function:

$$Q_n(\lambda) = \sum_{t=1}^{n}[Y_t - E(Y_t|Y_{t-1})]^2 = \sum_{t=1}^{n}(Y_t - \lambda)^2$$

Equation 5-4

with respect to λ for a sample of $\{Y_1, Y_2, \dots, Y_n\}$.

$$\hat{\lambda} = \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 5-5

The likelihood function of a sample of n observations of an INARMA(0,0) process can be written as:

$$L(\lambda) = \prod_{t=1}^{n} \frac{e^{-\lambda}\lambda^{Y_t}}{(Y_t)!}$$

Equation 5-6

The ML estimator of λ can be obtained by maximizing the log of the likelihood function in Equation 5-6. It can be seen that this results in the same estimator as that of CLS (Equation 5-5).

5.3 Estimation in an INAR(1) Model

5.3.1 YW for INAR(1)

The Yule-Walker estimator for α in an INAR(1) model was found by Al-Osh and Alzaid (1987) to be as follows:

$$\hat{\alpha} = \frac{\sum_{t=0}^{n-1}(Y_t - \bar{Y})(Y_{t+1} - \bar{Y})}{\sum_{t=0}^{n}(Y_t - \bar{Y})^2}$$

Equation 5-7

where $\bar{Y}$ is the sample mean. Since Z_t is assumed to have a Poisson distribution with parameter λ, the estimator for λ is:

$$\hat{\lambda}_{AA} = \frac{\sum_{t=1}^{n} \hat{Z}_t}{n}$$

Equation 5-8

where $\hat{Z}_t = Y_t - \hat{\alpha} Y_{t-1}$.

Jung and Tremayne (2006b) propose the same estimator for α, but a slightly different estimator for λ, which is indicated by $\hat{\lambda}_{JT}$. They argue that, as the first order moment of the INAR(1) model is $E(Y_t) = \frac{\lambda}{1-\alpha}$, λ can be estimated from:

$$\hat{\lambda}_{JT} = (1 - \hat{\alpha}) \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 5-9

In this thesis we use the Equation 5-7 and Equation 5-9 to obtain the YW estimates of an INAR(1) process. This is because the Equation 5-9 is based on observed data and not estimates of the innovations as in Equation 5-8.

5.3.2 CLS for INAR(1)

It can be easily seen that in the INAR(1) model, Y_t given Y_{t-1} is still a random variable due to the definition of the thinning operation. The conditional mean of Y_t given Y_{t-1}, which is the best one-step-ahead predictor (Brännäs and Hall, 2001), is:

$$E(Y_t | Y_{t-1}) = \alpha Y_{t-1} + \lambda = g(\boldsymbol{\theta}, Y_{t-1})$$

Equation 5-10

where $\boldsymbol{\theta} = (\alpha, \lambda)'$ is the vector of parameters to be estimated. Al-Osh and Alzaid (1987) employ a procedure developed by Klimko and Nelson (1978) and derive the estimators for α and λ as follows:

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} Y_t Y_{t-1} - (\sum_{t=1}^{n} Y_t \sum_{t=1}^{n} Y_{t-1})/n}{\sum_{t=1}^{n} Y_{t-1}^2 - (\sum_{t=1}^{n} Y_{t-1})^2/n}$$

Equation 5-11

and

$$\hat{\lambda} = \left[\sum_{t=1}^{n} Y_t - \hat{\alpha} \sum_{t=1}^{n} Y_{t-1}\right] / n$$

Equation 5-12

It can be easily verified that g, $\partial g/\partial \alpha$, $\partial g/\partial \lambda$, $\partial^2 g/\partial \alpha \partial \lambda$ satisfy the regularity conditions proposed by Klimko and Nelson (1978). It follows that the CLS estimators are strongly consistent and asymptotically normally distributed as:

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\big) \sim N(\mathbf{0}, V^{-1}WV)$$

where $\boldsymbol{\theta}^0 = (\alpha^0, \lambda^0)'$ denotes the true values of the parameters and:

$$V_{ij} = E\left[\frac{\partial g(\boldsymbol{\theta}^0, Y_{t-1})}{\partial \theta_i} . \frac{\partial g(\boldsymbol{\theta}^0, Y_{t-1})}{\partial \theta_j}\right] \qquad i, j = 1,2$$

$$W_{ij} = E\left[u_t^2(\boldsymbol{\theta}^0) \frac{\partial g(\boldsymbol{\theta}^0, Y_{t-1})}{\partial \theta_i} . \frac{\partial g(\boldsymbol{\theta}^0, Y_{t-1})}{\partial \theta_j}\right] \qquad i, j = 1,2$$

with $u_t(\boldsymbol{\theta}^0) = Y_t - g(\boldsymbol{\theta}^0, Y_{t-1})$.

Freeland and McCabe (2005) show that the distributions of the CLS and YW estimators of a PoINAR(1) process are asymptotically equivalent.

5.3.3 CML for INAR(1)

The Conditional Maximum Likelihood (CML) and Maximum Likelihood (ML) estimators for the PoINAR(1) process are provided by Al-Osh and Alzaid (1987). The likelihood function of a sample of n observations from an INAR(1) process can be written as:

$$L(\alpha, \lambda) = P(Y_1) \prod_{t=2}^{n} P(Y_t | Y_{t-1})$$

Equation 5-13

where $P(Y_t | Y_{t-1})$ is given by:

$$P(Y_t|Y_{t-1}) = \sum_{i=0}^{\min(Y_{t-1},Y_t)} \binom{Y_{t-1}}{i} \alpha^i (1-\alpha)^{Y_{t-1}-i} \frac{e^{-\lambda}\lambda^{Y_t-i}}{(Y_t-i)!}$$

Equation 5-14

Because the marginal distribution of the PoINAR(1) process is Poisson with mean $\lambda/(1-\alpha)$, the unconditional likelihood function is:

$$L(\alpha,\lambda) = \frac{e^{-\lambda/(1-\alpha)}[\frac{\lambda}{1-\alpha}]^{Y_1}}{(Y_1)!} \prod_{t=2}^{n} \left[\sum_{i=0}^{\min(Y_{t-1},Y_t)} \binom{Y_{t-1}}{i} \alpha^i (1-\alpha)^{Y_{t-1}-i} \frac{e^{-\lambda}\lambda^{Y_t-i}}{(Y_t-i)!} \right]$$

Equation 5-15

In order to find the conditional maximum likelihood estimation (CML), Y_1 is assumed to be given and the conditional likelihood function is reduced to:

$$L(\alpha,\lambda) = \prod_{t=2}^{n} \left[\sum_{i=0}^{\min(Y_{t-1},Y_t)} \binom{Y_{t-1}}{i} \alpha^i (1-\alpha)^{Y_{t-1}-i} \frac{e^{-\lambda}\lambda^{Y_t-i}}{(Y_t-i)!} \right]$$

Equation 5-16

Then, the unconditional and conditional maximum likelihood estimators can be derived by maximizing the logarithm of the likelihood functions of Equation 5-15 and Equation 5-16, respectively.

Al-Osh and Alzaid (1987) used the procedure of Sprott (1983) to eliminate one of the parameters in the derivatives of the log-likelihood function.

$$\frac{\partial \log[L(\alpha,\lambda)]}{\partial \lambda} = \sum_{t=2}^{n} H(t) - (n-1) = 0$$

Equation 5-17

$$\frac{\partial \log[L(\alpha,\lambda)]}{\partial \alpha} = \sum_{t=2}^{n} \frac{(Y_t - \alpha Y_{t-1}) - \lambda H(t)}{\alpha(1-\alpha)} = 0$$

Equation 5-18

where $H(t) = P(Y_t - 1)/P(Y_t)$. The Equation 5-18 results in:

$$\hat{\lambda} = \frac{\sum_{t=2}^{n} Y_t - \alpha \sum_{t=2}^{n} Y_{t-1}}{n-1}$$

Equation 5-19

The Equation 5-19 can then be used in Equation 5-17 to find $\hat{\alpha}$. The ML estimators of α and λ have the following asymptotic distribution:

$$\sqrt{n}\begin{pmatrix}\hat{\alpha}-\alpha\\ \hat{\lambda}-\lambda\end{pmatrix}\sim N(\mathbf{0},i^{-1})$$

where the matrix i is the Fisher information (Bu, 2006).

Al-Osh and Alzaid (1987) compare the performance of YW, CLS and CML estimates in terms of bias and MSE in a simulation study. Their results suggest that, in general, CML is worth the extra effort because it has the least bias and MSE of all. However, for small sample sizes $(n \le 75)$ and small autoregressive parameter $(\alpha = 0.1)$, because the sample contains many zero values, CML is not as good as YW in terms of bias and MSE. It is worth mentioning that their study only compares the accuracy of estimates and not their impact on forecast accuracy, which is done in this PhD thesis (see section 8.4).

5.3.4 Conditional GMM for INAR(1)

Brännäs (1994) uses the conditional GMM estimation method of Hansen (1982) to estimate the parameters of a PoINAR(1) process. It is called a conditional GMM since the moment restrictions used are conditional. The GMM estimator is based on minimization of the function:

$$q = \mathbf{m}(\boldsymbol{\theta})'\widehat{\mathbf{W}}^{-1}\mathbf{m}(\boldsymbol{\theta})$$

Equation 5-20

where $\boldsymbol{\theta} = (\alpha,\lambda)'$ is the vector of the unknown parameters to be estimated and $\mathbf{m}(\boldsymbol{\theta})$ is the vector of moment restrictions. When $\widehat{\mathbf{W}}$ is the asymptotic covariance of $\mathbf{m}(\boldsymbol{\theta})$, the GMM estimator is efficient. q is first minimized based on using an identity matrix $\mathbf{I}$ for $\widehat{\mathbf{W}}$. Then the estimates $\widehat{\boldsymbol{\theta}}$ are used to from $\widehat{\mathbf{W}}$.

The moment restrictions used are:

$$\frac{1}{n}\sum_{t=2}^{n} e_t = 0$$

$$\frac{1}{n}\sum_{t=2}^{n} E(Y_{t-1}e_t) = 0$$

where e_t is the one-step ahead prediction error, $e_t = Y_t - E(\, Y_t | Y_{t-1})$.

Although the above moments are unconditional, they are equal to the conditional ones (Brännäs, 1994). It can be seen that when $\mathbf{W} = \mathbf{I}_2$, the GMM and the CLS are the same.

Brännäs (1994) then compares the performance of CLS, ML, and GMM. The results show that for large values of α, the GMM estimates have smaller MSE than CLS. In general, for large values of α, $\hat{\alpha}_{\mathrm{GMM}}$ is close to $\hat{\alpha}_{\mathrm{ML}}$ in terms of MSE. But for small values of α, $\hat{\alpha}_{\mathrm{ML}}$ outperforms $\hat{\alpha}_{\mathrm{GMM}}$. Moreover, $\hat{\lambda}_{\mathrm{ML}}$ always has smaller MSE than $\hat{\lambda}_{\mathrm{GMM}}$, although when α increases this difference decreases. The results do not show any conclusive advantage by using GMM over ML in terms of bias.

To conclude, we do not see any benefit in using GMM for an INAR(1) process, considering the fact that it does not outperform the maximum likelihood method, which is used in this research, in terms of MSE.

5.4 Estimation in an INAR(*p*) Model

5.4.1 YW for INAR(*p*)

Du and Li (1991) generalize the Yule-Walker estimation method to estimate the parameters of an INAR(*p*) process. The Yule-Walker equations are:

$$\Gamma\boldsymbol{\alpha} = \boldsymbol{\rho}$$

Equation 5-21

where $\Gamma = [\rho_{|i-j|}]_{P\times P}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)'$, and $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_p)'$. Replacing the theoretical autocorrelations ρ_k by the sample autocorrelations r_k results in the YW estimate of $\boldsymbol{\alpha}$.

Then, λ can be estimated from the expected value of the INAR(*p*) process:

$$\hat{\lambda} = (1 - \hat{\alpha}_1 - \cdots - \hat{\alpha}_p)\bar{Y}$$

Equation 5-22

These estimators are strongly consistent and asymptotically normally distributed (Silva and Silva, 2006). Also, for a large number of observations, YW estimators are very close to the CLS estimators (Du and Li, 1991).

5.4.2 CLS for INAR(p)

The CLS estimators of an INAR(p) process are also derived by Du and Li (1991). The conditional expected value of the process is:

$$E(Y_t|Y_{t-1}, \dots, Y_2, Y_1) = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \lambda$$

Equation 5-23

The least squares criterion to be minimized is then:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=p+1}^{n} [Y_t - E(Y_t|Y_{t-1}, \dots, Y_2, Y_1)]^2$$

Equation 5-24

where $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \dots, \alpha_p, \lambda)$ is the vector of parameters to be estimated. $\hat{\boldsymbol{\theta}}$ can be found by setting the partial derivatives to zero.

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \alpha_i} = 0 \qquad (i = 1,2, \dots, p)$$

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \lambda} = 0$$

Du and Li (1991) suggest that, for large samples, the CLS estimators for INAR(p) process are very close to Yule-Walker estimators. They also are strongly consistent and asymptotically normal.

5.4.3 CML for INAR(p)

Bu et al. (2008) study the maximum likelihood estimators of a general INAR(p)

process. The conditional likelihood function of an INAR(p) process is:

$$L(\alpha_1, \alpha_2, \dots, \alpha_p, \lambda) = \prod_{t=p+1}^{n} P(Y_t | Y_{t-1}, \dots, Y_{t-p})$$

Equation 5-25

where the conditional probabilities are:

$$P(Y_t | Y_{t-1}, \dots, Y_{t-p}) = \sum_{i_1=0}^{\min(Y_{t-1}, Y_t)} \binom{Y_{t-1}}{i_1} \alpha_1^{i_1} (1-\alpha_1)^{Y_{t-1}-i_1} \times \sum_{i_2=0}^{\min(Y_{t-2}, Y_t - i_1)} \binom{Y_{t-2}}{i_2} \alpha_2^{i_1} (1-\alpha_2)^{Y_{t-2}-i_2} \times \dots \times \sum_{i_p=0}^{\min[Y_{t-p}, Y_t - (i_1 + \dots + i_{p-1})]} \binom{Y_{t-p}}{i_p} \alpha_p^{i_p} (1-\alpha_p)^{Y_{t-p}-i_p} \frac{e^{-\lambda} \lambda^{Y_t - (i_1 + \dots + i_{p-1})}}{[Y_t - (i_1 + \dots + i_{p-1})]!}$$

Equation 5-26

The asymptotic distribution of the maximum likelihood estimator is Normal:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(\mathbf{0}, i^{-1})$$

where $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \lambda)'$ is the parameter vector and the matrix i is the Fisher information.

Bu et al. (2008) investigate the asymptotic benefit of ML over CLS for a PoINAR(2) process using the asymptotic relative efficiency (ARE) ratio between the two estimators, i.e. the ratio of their asymptotic variances. Their results show that for persistent processes (high values of α_1, α_2 or both), the ML estimates are asymptotically more efficient than the CLS estimates. For low values of α_1 and α_2 and also for high values of α_2 and low values of α_1, the benefit of ML over CLS is slight.

They also compare the performance of ML and CLS in a simulation study for $n = 100$ and $n = 500$. The results suggest that there is a gain in terms of MSE in using ML for larger samples. For smaller samples, when α_1 and α_2 are small, CLS has lower MSE than ML.

5.5 Estimation in an INMA(1) Model

5.5.1 YW for INMA(1)

The Yule-Walker estimator for β in a PoINMA(1) process ($Y_t = Z_t + \beta \circ Z_{t-1}$) is as follows:

$$\hat{\beta} = \frac{r_1}{1 - r_1}$$

Equation 5-27

where r_1 is the lag one sample autocorrelation given by the Equation 5-1. Then, λ can be estimated from the expected value of the process:

$$\hat{\lambda} = \frac{1}{1 + \hat{\beta}} \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 5-28

Note that λ can also be estimated form the unconditional variance of the INMA(1) (Equation 3-28), but the estimator based on the unconditional mean has smaller variance (Brännäs and Hall, 2001).

5.5.2 CLS for INMA(1)

The conditional expected value of Y_t given Y_{t-1} for an INMA(1) process is given by:

$$E(Y_t|Y_{t-1}) = \beta Z_{t-1} + \lambda$$

Equation 5-29

The prediction error is:

$$e_t = Y_t - \beta Z_{t-1} - \lambda$$

Equation 5-30

The CLS estimates of β and λ can then obtained by minimizing the following function:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [Y_t - (\beta Z_{t-1} + \lambda)]^2$$

Equation 5-31

with respect to $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\beta, \lambda)'$ is the parameter vector to be estimated. The CLS estimates for β and λ are:

$$\hat{\beta} = \frac{\sum_{t=1}^{n} Y_t Z_{t-1} - (\sum_{t=1}^{n} Y_t \sum_{t=1}^{n} Z_{t-1})/n}{\sum_{t=1}^{n} Z_{t-1}^2 - (\sum_{t=1}^{n} Z_{t-1})^2/n}$$

Equation 5-32

$$\hat{\lambda} = \left[\sum_{t=1}^{n} Y_t - \hat{\beta} \sum_{t=1}^{n} Z_{t-1}\right]/n$$

Equation 5-33

5.6 Estimation in an INMA(q) Model

5.6.1 YW for INMA(q)

The autocorrelation function of an INMA(q) process of $Y_t = Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-\mathrm{q}}$ is given by the Equation 3-40. The ACF can be used to find the YW estimates of $(\beta_1, \dots, \beta_q)$.

Once these parameters have been estimated, λ can be estimated from the expected value of the process:

$$\hat{\lambda} = \frac{1}{1 + \sum_{j=1}^{q} \beta_j} \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 5-34

When the order of the INMA model increases, the equations to be solved become more complex. This is shown in Table 5-2.

Table 5-2 The relationship between the order of the model and the type of YW equation

Model	Equation
INMA(1)	Linear
INMA(2)	quadratic (2)
INMA(3)	quartic (4)
INMA(4)	sextic (6)

The results of Table 5-2 can be found by direct expansion of expressions in the Yule-Walker equations. As can be seen from the above table, for INMA processes with order higher than two, it becomes more complex to find the estimators and although such equations can be solved numerically, it is computationally expensive to find the Yule-Walker estimators for such processes.

5.6.2 CLS for INMA(q)

The conditional first moment of the INMA(q) process is:

$$E(Y_t|Y_{t-1}) = \lambda + \sum_{j=1}^{q} \beta_j Z_{t-\mathrm{j}}$$

Equation 5-35

Hence, the forecast error is:

$$e_t = Y_t - \lambda - \sum_{j=1}^{q} \beta_j Z_{t-j}$$

Equation 5-36

The least squares criterion is then:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=q+1}^{n} \left[Y_t - \lambda - \sum_{j=1}^{q} \beta_j Z_{t-j} \right]^2$$

Equation 5-37

The corresponding parameter vector $\boldsymbol{\theta} = (\beta_1, \dots, \beta_q, \lambda)'$ for the minimum $Q_n(\boldsymbol{\theta})$ can be obtained.

5.6.3 GMM based on Probability Generation Functions for INMA(q)

In this estimation method, the probability generation functions (pgf) $\phi(z)$ and $\Phi_k(z_1, z_2)$ are evaluated at any z. Based on the law of large numbers:

$$\frac{\sum_{t=1}^{n}(1-z_p)^{Y_t}}{n} \xrightarrow{p} \phi(z_p)$$

$$\frac{\sum_{t=k+1}^{n}(1-z_p)^{Y_t}(1-z_n)^{Y_{t-k}}}{n-k} \xrightarrow{p} \Phi_k(z_p, z_n)$$

where $\xrightarrow{p}$ denotes convergence in probability. Therefore, the moment conditions are formed as:

$$m_{1p} = \frac{\sum_{t=1}^{n}(1-z_p)^{Y_t}}{n} - \phi(z_p)$$

$$m_{2k,pn} = \frac{\sum_{t=k+1}^{n}(1-z_p)^{Y_t}(1-z_n)^{Y_{t-k}}}{n-k} - \Phi_k(z_p, z_n)$$

The GMM criterion to be minimized is then:

$$h = \mathbf{m}'\mathbf{W}^{-1}\mathbf{m}$$

where **m** is the vector of moment restrictions and **W** is the covariance matrix of **m**. Similar to section 5.3.4, first it is assumed that **W** is equal to the identity matrix. This results in a consistent and asymptotically normal estimator. However, if **W** is known or a consistent estimator of **W** is used, the GMM estimators would be more efficient than if **W** is equal to **I**.

For a PoINMA(1) process, GMM estimators provided by Brännäs and Hall (2001) are as follows:

$$\hat{\beta} = \frac{[\ln(e) - \ln(v)]z_1 - \ln(e)z_2}{[\ln(e) - \ln(v)]z_1 - \ln(e)z_2(1-z_1)}$$

Equation 5-38

$$\hat{\lambda} = -\frac{\ln(e)}{z_1(1+\hat{\beta})}$$

Equation 5-39

where e and v are the sample moments corresponding to $\phi(z_1)$ and $\Phi_1(z_1, z_2)$, respectively.

Brännäs and Hall (2001) compare the performance of YW, CLS and GMM for an

INMA(2) process. The results for the case that $\beta_1 = \beta_2 = 0.9$ (note that the process is not invertible) suggest that GMM has the smallest bias and MSE among all estimators except for small sample sizes where CLS is the best. For all β_2 values except $\beta_2 = 0.9$, CLS is better than GMM in terms of MSE.

Although they mention that for other values of β_1 and β_2 GMM has smaller bias and MSE in most cases, they did not present any results for further comparisons.

Brännäs and Hall (2001) explain that the performance of the GMM estimator depends highly on the selection of z-values. Finally, because of the better overall performance and its simplicity, they advocate the use of CLS instead of GMM.

5.7 Estimation in an INARMA(1,1) Model

In this section, the YW and CLS estimators for the parameters of an INARMA(1,1) process, derived in this PhD research, are presented.

5.7.1 YW for INARMA(1,1)

The ACF of an INARMA(1,1) process of $Y_t = \alpha \circ Y_{t-1} + Z_t + \beta \circ Z_{t-1}$ is given by Equation 3-48. When the distribution of the innovations $\{Z_t\}$ is Poisson, the ACF is:

$$\rho_k = \begin{cases} \dfrac{\alpha + \beta + \alpha\beta + \alpha^2 + 2\alpha^2\beta}{1 + \alpha + \beta + 3\alpha\beta} & \text{for } k = 1 \\ \alpha\rho_{k-1} & \text{for } k > 1 \end{cases}$$

Equation 5-40

Hence, α and β can be estimated from:

$$\hat{\alpha} = \frac{r_2}{r_1} = \frac{\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}$$

Equation 5-41

$$\hat{\beta} = \frac{(1 + \hat{\alpha})(\hat{\alpha} - r_1)}{r_1(1 + 3\hat{\alpha}) - 1 - \hat{\alpha} - 2\hat{\alpha}^2}$$

Equation 5-42

Then, λ can be estimated using the expected value of an INARMA(1,1) process:

$$\hat{\lambda} = \frac{1-\hat{\alpha}}{1+\hat{\beta}} \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 5-43

5.7.2 CLS for INARMA(1,1)

The conditional expected value of an INARMA(1,1) process is:

$$E(Y_t|Y_{t-1}) = \alpha Y_{t-1} + \lambda + \beta Z_{t-1}$$

Equation 5-44

The conditional least squares criterion is therefore:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})]^2$$

Equation 5-45

with $\boldsymbol{\theta} = (\alpha, \beta, \lambda)'$ is the parameter vector to be estimated. The estimators for α, β, and λ can then be obtained by minimizing the above function with respect to $\boldsymbol{\theta}$.

$$\hat{\alpha} = \frac{\left\{\begin{matrix} n^2 \sum Y_t Y_{t-1} \sum Z_{t-1}^2 - n \sum Y_t \sum Y_{t-1} \sum Z_{t-1}^2 - n \sum Y_t Y_{t-1} (\sum Z_{t-1})^2 + n \sum Y_{t-1} \sum Z_{t-1} \sum Y_t Z_{t-1} \\ -n^2 \sum Y_t Z_{t-1} \sum Y_{t-1} Z_{t-1} + n \sum Y_t \sum Z_{t-1} \sum Y_{t-1} Z_{t-1} \end{matrix}\right\}}{n^2 \sum Y_{t-1}^2 \sum Z_{t-1}^2 - n(\sum Y_{t-1})^2 \sum Z_{t-1}^2 - n \sum Y_{t-1}^2 (\sum Z_{t-1})^2 + 2n \sum Y_{t-1} \sum Z_{t-1} \sum Y_{t-1} Z_{t-1} - n^2 (\sum Y_{t-1} Z_{t-1})^2}$$

Equation 5-46

$$\hat{\beta} = \frac{n \sum Y_t Z_{t-1} - n\hat{\alpha} \sum Y_{t-1} Z_{t-1} - \sum Y_t \sum Z_{t-1} + \hat{\alpha} \sum Y_{t-1} \sum Z_{t-1}}{n \sum Z_{t-1}^2 - (\sum Z_{t-1})^2}$$

Equation 5-47

$$\hat{\lambda} = \frac{\sum Y_t - \hat{\alpha} \sum Y_{t-1} - \hat{\beta} \sum Z_{t-1}}{n}$$

Equation 5-48

where all the summations are from 1 to n (see Appendix 5.A for the proof).

5.8 YW Estimators of an INARMA(2,2) Model

The YW estimators of an INARMA(2,2) model can be obtained from the ACF of an INARMA(p,q) model derived in chapter 3. The INARMA(2,2) model has the form:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t + \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2}$$

Equation 5-49

From the Equation 3-52, the variance of the above process, when the innovations are Poisson distributed, the variance of an INARMA(2,2) process can be found as (see Appendix 5.B):

$$\mathrm{var}(Y_t) = \frac{\lambda}{(1-\alpha_1^2-\alpha_2^2)(1-\alpha_2)-2\alpha_1^2\alpha_2} \times \left\{(1-\alpha_2)\left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}(\alpha_1-\alpha_1^2+\alpha_2-\alpha_2^2)+1+\beta_1+\beta_2+2\alpha_1\beta_1 + 2\alpha_2\beta_2\right]+2\alpha_1\alpha_2\beta_1+2\alpha_1^2\beta_2+2\alpha_1\beta_1\beta_2\right\}$$

Equation 5-50

The autocorrelation function of an INARMA(2,2) process can be found from the Equation 3-55 as:

$$\rho_k = \begin{cases} \dfrac{\alpha_1\gamma_0+\alpha_2\gamma_1+\beta_1\lambda+\beta_2(\alpha_1+\beta_1)\lambda}{\gamma_0} & for\ k=1 \\ \dfrac{\alpha_1\gamma_1+\alpha_2\gamma_0+\beta_2\lambda}{\gamma_0} & for\ k=2 \\ \alpha_1\rho_{k-1}+\alpha_2\rho_{k-2} & for\ k>2 \end{cases}$$

Equation 5-51

Based on the Equation 5-51, the autocorrelation of lags one to four can be used to estimate α_1, α_2, β_1, and β_2.

$$\rho_1 = \alpha_1 + \alpha_2\rho_1 + \frac{\beta_1\lambda+\beta_2(\alpha_1+\beta_1)\lambda}{\gamma_0}$$

Equation 5-52

$$\rho_2 = \alpha_1\rho_1 + \alpha_2 + \frac{\beta_2\lambda}{\gamma_0}$$

Equation 5-53

$$\rho_3 = \alpha_1\rho_2 + \alpha_2\rho_1$$

Equation 5-54

$$\rho_4 = \alpha_1 \rho_3 + \alpha_2 \rho_2$$

Equation 5-55

The last two equations can be used to find $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

$$\hat{\alpha}_1 = \frac{r_1 r_4 - r_2 r_3}{r_1 r_3 - r_2^2}$$

Equation 5-56

$$\hat{\alpha}_2 = \frac{r_3 - \hat{\alpha}_1 r_2}{r_1}$$

Equation 5-57

where r_k is the sample autocorrelation at lag k given by the Equation 5-1. Then, β_1 and β_2 can be found from Equation 5-52 and Equation 5-53.

Finally, the expected value of the process can be used to estimate λ.

$$\hat{\lambda} = \left(\frac{1 - \hat{\alpha}_1 - \hat{\alpha}_2}{1 + \hat{\beta}_1 + \hat{\beta}_2}\right) \bar{Y}$$

Equation 5-58

where $\bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n}$.

5.9 Conclusions

Different methods for estimating the parameters of INARMA models provided in the literature have been reviewed in this chapter. This includes YW, CLS, CML, and GMM. Not all these methods have been developed for all INARMA models. For example, the maximum likelihood function and therefore the CML estimators have been developed only for INAR(p) models.

The performance of these estimators has been compared in some studies (Al-Osh and Alzaid, 1987; Brännäs, 1994; Brännäs, 1995; Brännäs and Hall, 2001; Bu et al., 2008). The results generally suggest that the CML is worth the extra effort especially for high values of α (and α_1 for an INAR(2) process) and reasonably large samples. For lower values of α and smaller samples, the CLS has lower MSE than ML.

The YW and CLS estimates for an INARMA(1,1) process are provided for the first time. The YW results are based on the ACF of an INARMA(p,q) model of Equation 3-55. These results along with YW, CLS, CML for INAR(1) process and YW and CLS for INMA(1) process will be used in simulation and empirical analyses of this PhD research (chapters 8 and 9).

Finding the ACF of an INARMA(p,q) process in chapter 3 has enabled us to derive the YW estimators for these processes. As a further example, these estimators are obtained for an INARMA(2,2) process.

We have decided not to follow the GMM method. This is because, for an INAR(1) process, it does not outperform the maximum likelihood estimator, which is covered by this study, in terms of MSE. For the INMA(1) process, no comparison has been done in the literature and the one that compared CLS and GMM for INMA(2) (Brännäs and Hall, 2001) does not provide results for all parameter sets.

It is worth mentioning that all of the above-mentioned studies only compared the bias and MSE of the estimates and not their impact on forecast accuracy. This will be done in our simulation experiment (see section 8.4).

Chapter 6 FORECASTING IN INARMA MODELS

6.1 Introduction

Having discussed some stochastic properties of INARMA models, and identification and estimation of the parameters of these models, we now investigate how these models can be used in forecasting future values of an observed time series. This section is organized as follows. The minimum mean square error (MMSE) forecasts for INAR(p), INMA(q) and INARMA(p,q) processes are reviewed in section 6.2. The lead time aggregation and forecasting of INARMA processes is then discussed in section 6.3. Results on lead time aggregation and forecasting for INAR(1), INMA(1), and INARMA(1,1) processes are presented. The conclusions are given in section 6.4.

6.2 MMSE Forecasts

The most common forecasting procedure discussed in the time series literature is using the conditional expectation (Freeland and McCabe, 2004b). The main advantage of this method, apart from being simple, is that it produces forecasts with minimum mean square error (MMSE).

Freeland and McCabe (2004b) argue that this method does not produce coherent forecasts for INARMA models. Coherency means that forecasts should comply with time series restrictions, in this case being integers. Freeland and McCabe (2004b) suggest the median of the distribution and use the h-step-ahead conditional distribution to produce coherent forecasts for the PoINAR(1) model. Bu and McCabe (2008) present a procedure to produce h-step-ahead distribution forecasts for the PoINAR(p) process using the transition probability function of the process. Jung and Tremayne (2006b) introduce a Monte Carlo procedure to estimate the h-step-ahead forecast distribution for INAR(1) and INAR(2) processes.

This PhD research tries to apply INARMA models for intermittent demand forecasting. We are especially interested in comparing the accuracy of forecasts produced by INARMA methods to non-optimal smoothing-based methods (the last research question, p.7). For this reason, we compare the point forecasts of all methods using the accuracy measures suggested in section 2.4.3 (including MSE). We focus on the conditional expectation since it provides the MMSE forecasts for INARMA methods. It will be discussed in chapter 10 that forecasting the whole distribution can be considered as a further research avenue.

6.2.1 MMSE Forecasts for an INAR(p) Model

Minimum mean square error (MMSE) forecasts are used to find $\hat{Y}_{T+h}$, $h = 1,2,\dots,H$ of the process Y_t based on the observed series of $\{Y_1,\dots,Y_T\}$. The MMSE forecast of the process is given by:

$$\hat{Y}_{T+h} = E(Y_{T+h}|Y_T,\dots,Y_1)$$

Equation 6-1

As indicated by its title, this method yields forecasts with minimum MSE. For an INAR(p) model of Equation 3-19, we have:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \alpha_2 Y_{T+h-2} + \cdots + \alpha_p Y_{T+h-p} + \lambda$$

Equation 6-2

where the Y values on the RHS of Equation 6-2 may be either actual or forecast values (Du and Li, 1991; Jung and Tremayne, 2006b). This is shown in Figure 6-1 for the case where $h \leq p$.

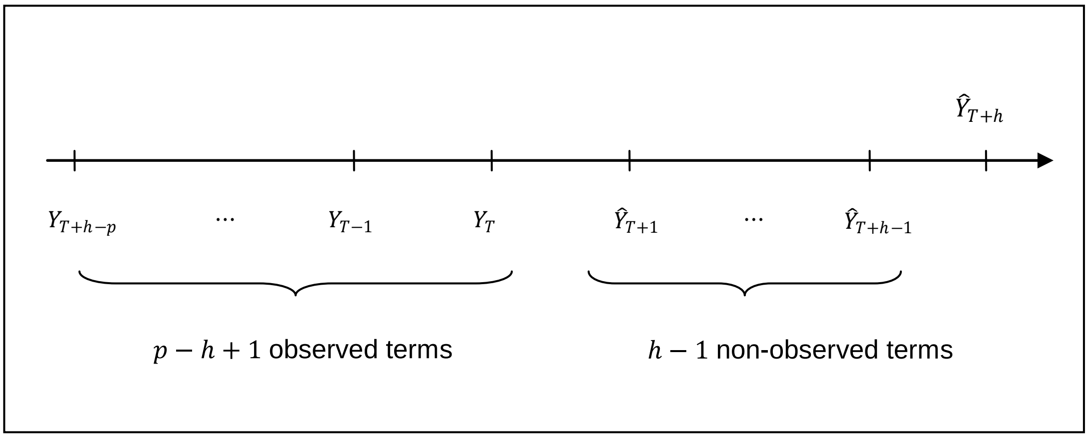

Figure 6-1 *h*-step-ahead forecast for an INAR(p) model when $h \leq p$

This is called using a single model for all horizons. For example, for an INAR(2) process, the *h*-step ahead forecast is given by:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \alpha_2 Y_{T+h-2} + \lambda$$

Equation 6-3

This implies that for large h, the forecasts converge to the unconditional mean of the INAR(2) process that is:

$$\hat{Y}_{T+h} \to \frac{\lambda}{1 - \alpha_1 - \alpha_2}$$

Some authors suggested that using different models for different horizons can improve forecast accuracy (Cox, 1961; Tiao and Xu, 1993; Kang, 2003). For an AR(p) model, this is:

$$Y_{t+h} = \alpha_{1,h} Y_t + \alpha_{2,h} Y_{t-1} + \cdots + \alpha_{p(h),h} Y_{t-p(h)+1} + \varepsilon_{t,h}$$

Equation 6-4

It can be seen from Equation 6-4 that even the order of the AR model depends on the forecast horizon h.

6.2.2 MMSE Forecasts for an INMA(q) Model

As studied by Brännäs and Hall (2001), for an INMA(q) process of Equation 3-36, the MMSE one-step ahead forecast can be obtained from:

$$\hat{Y}_{T+1} = \beta_1 Z_T + \beta_2 Z_{T-1} + \cdots + \beta_q Z_{T-q+1} + \lambda$$

Equation 6-5

The forecast error variance is:

$$\text{var}(e_t) = \lambda \left[1 + \sum_{j=1}^{q} \beta_j (1 - \beta_j) \right]$$

Note that when X is a random variable, $\text{var}(\alpha \circ X) = \alpha^2 \text{var}(X) + \alpha(1-\alpha)E(X)$, but when X is given as in the above case, $\text{var}(\alpha \circ X) = \alpha(1-\alpha)E(X)$.

The h-step ahead forecast when $h \leq q$ is given by:

$$\hat{Y}_{T+h} = \beta_h Z_T + \cdots + \beta_q Z_{T+h-q} + \lambda(1 + \beta_1 + \cdots + \beta_{h-1})$$

Equation 6-6

This is shown in Figure 6-2. In the above equation, the Z values on the RHS can be estimated from the previous estimated Zs and observed Ys based on Equation 3-36.

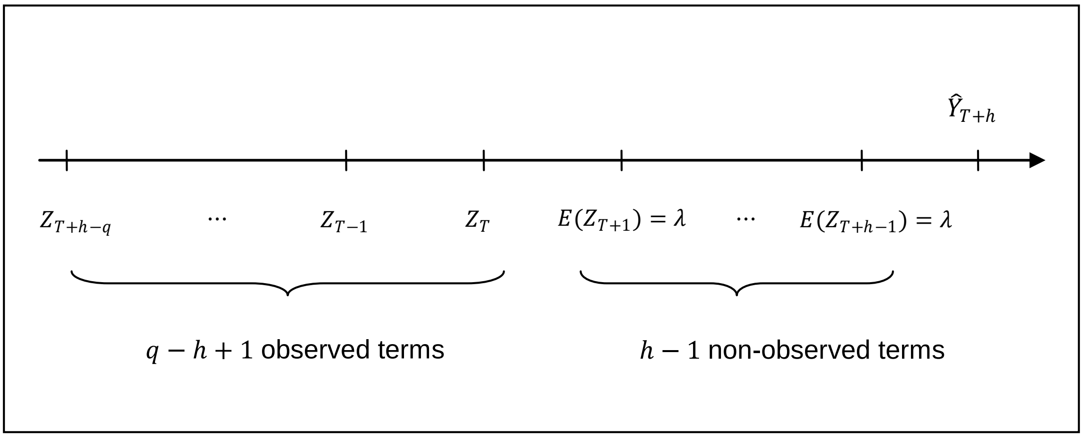

Figure 6-2 h-step-ahead forecast for an INMA(q) model when $h \leq q$

The forecast error variance for $h \leq q$ is:

$$\text{var}(e_t) = \lambda\left[1 + \sum_{j=1}^{h-1} \beta_j + \sum_{j=h}^{q} \beta_j (1 - \beta_j)\right]$$

When $h > q$, the h-step ahead forecast becomes:

$$\hat{Y}_{T+h} = \lambda\left[1 + \sum_{j=1}^{q} \beta_j\right]$$

Equation 6-7

with the forecast error variance of $\text{var}(e_t) = \lambda\left[1 + \sum_{j=1}^{q} \beta_j\right]$.

6.2.3 MMSE Forecasts for an INARMA(*p*,*q*) Model

The above results can be generalized for an INARMA(*p*,*q*) process. The MMSE one-step-ahead forecast is then:

$$\hat{Y}_{T+1} = \alpha_1 Y_T + \cdots + \alpha_p Y_{T-p+1} + \lambda + \beta_1 Z_T + \cdots + \beta_q Z_{T-q+1}$$

Equation 6-8

The h-step ahead forecast when $h \leq q$ will be:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \cdots + \alpha_p Y_{T+h-p} + \lambda + \beta_h Z_T + \cdots + \beta_q Z_{T+h-q} + \lambda(\beta_1 + \cdots + \beta_{h-1})$$

Equation 6-9

where Y values on the RHS of Equation 6-9 may be either actual or forecast values. When $h > q$, the h-step ahead forecast becomes:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \cdots + \alpha_p Y_{T+h-p} + \lambda \sum_{j=0}^{q} \beta_j$$

Equation 6-10

where again Y values on the RHS of the above equation may be either actual or forecast values and $\beta_0 = 1$.

6.3 Forecasting over Lead Time

In this section, forecasting over a lead time is discussed. Lead time forecasting has applications in many areas, particularly in an inventory management context, where forecasts are needed over the period that it takes from placing an order to receiving it from the supplier.

It can be easily seen that for an INARMA(0,0) process, the lead time aggregated process is:

$$\sum_{j=1}^{l+1} Y_{t+j} = Z_{t+1} + Z_{t+2} + \cdots + Z_{t+l+1} = \sum_{j=1}^{l+1} Z_{t+j}$$

Therefore, the conditional expected value and variance of the above equation are:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = \mathrm{var}\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = (l+1)\lambda$$

which is expected as the aggregated process is the sum of $(l+1)$ independent Poisson random variables which is in fact a Poisson variable with parameter $(l+1)\lambda$.

This section is organized as follows. First, the results of over-lead time aggregation and forecasting of INAR(1) and INMA(1) processes are presented. These results, along with similar results for INAR(2) and INARMA(1,2) processes in Appendices 6.A and 6.B, will then help us to find the over-lead time aggregation of the INARMA(p,q) process. The corresponding results for an INARMA(1,1) process are also provided, which will be used in chapters 8 and 9.

6.3.1 Lead Time Forecasting for an INAR(1) Model

For the INAR(1) process of $Y_t = \alpha \circ Y_{t-1} + Z_t$, the cumulative Y over lead time l is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1}$$

$$= (\alpha \circ Y_t + Z_{t+1}) + (\alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + Z_{t+2})$$
$$+ \cdots + (\alpha^{l+1} \circ Y_t + \alpha^l \circ Z_{t+1} + \alpha^{l-1} \circ Z_{t+2} + \cdots + Z_{t+l+1})$$

Equation 6-11

Because $\alpha \circ X + \beta \circ X \neq (\alpha + \beta) \circ X$, the above equation can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1} \sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1} \sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Z_{t+k_{ij}}$$

Equation 6-12

where n_j^1 is the number of Y_t terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in Equation 6-11, ψ_{ij}^1 is the corresponding coefficient for each Y_t, n_j^2 is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in Equation 6-11, and ψ_{ij}^2 is the corresponding coefficient for each $Z_{t+k_{ij}}$. All of these terms are explained below.

It can be seen that because the process is an integer autoregressive of order one, each of $\{Y_{t+j}\}_{j=1}^{l+1}$ yields only one Y_t in Equation 6-11; therefore, $n_j^1 = 1$. The corresponding coefficient for Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (say Y_{t+2}) is obtained from α thinned the coefficient of Y_t in the previous term (in this case Y_{t+1}). As a result, $\psi_{ij}^1 = \alpha^j$. These coefficients are shown in Table 6-1.

Table 6-1 Coefficients of Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INAR(1) model

$j = 1, i = 1$	$\psi_{11}^1 = \alpha$
$j = 2, i = 1$	$\psi_{12}^1 = \alpha^2$
$\vdots$	$\vdots$
$j = l + 1, i = 1$	$\psi_{1(l+1)}^1 = \alpha^{l+1}$

It can be seen from Equation 6-11 that due to the repeated substitution of Y_{t+j}, the number of $Z_{t+k_{ij}}$ increases in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. This number, shown by n_j^2, can be obtained from $n_{j-1}^2 + 1$. This means that each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (say Y_{t+2}) has one more Z compared to the previous one (which is Y_{t+1} in this case). The corresponding coefficient for each $Z_{t+k_{ij}}$, shown by ψ_{ij}^2, is α thinned the corresponding coefficient in the previous term ($\alpha \circ \psi_{i(j-1)}^2$). $t + k_{ij}$ is the subscript of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ and from Equation 6-11 it can be easily seen that k_{ij} is given by Equation 6-13. All of these terms are shown in Table 6-2.

$$k_{ij} = \begin{cases} k_{i(j-1)} & \text{for } 1 \le i \le n_{j-1}^2 \\ j & \text{for } n_{j-1}^2 < i \le n_j^2 \end{cases} \qquad \text{for } j = 1, \dots, l+1$$

Equation 6-13

Table 6-2 Coefficients of $Z_{t+k_{ij}}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INAR(1) model

$j = 1$ $i = 1, \dots, n_1^2$ where $n_1^2 = 1$	$\psi_{11}^2 = 1$	$k_{11} = 1$
$j = 2$ $i = 1, \dots, n_2^2$ where $n_2^2 = 2$	$\psi_{12}^2 = \alpha$ $\psi_{22}^2 = 1$	$k_{12} = 1$ $k_{22} = 2$
$\vdots$	$\vdots$	$\vdots$
$j = l+1$ $i = 1, \dots, n_{l+1}^2$ where $n_{l+1}^2 = l+1$	$\psi_{1(l+1)}^2 = \alpha^l$ $\vdots$ $\psi_{(l+1)(l+1)}^2 = 1$	$k_{1(l+1)} = 1$ $\vdots$ $k_{(l+1)(l+1)} = l+1$

Based on Equation 6-12, the conditional expected value of the aggregated process is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2\right)\lambda = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t$$
$$+ \left(\sum_{j=1}^{l+1}\sum_{i=1}^{j} \alpha^{i-1}\right)\lambda = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right]$$

Equation 6-14

which is the same as Equation 3-60.

Therefore at time T, when Y_T is observed, the lead time forecast can be obtained from:

$$E\left(\sum_{j=1}^{l+1} Y_{T+j} \mid Y_T\right) = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_T + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right]$$

Equation 6-15

6.3.2 Lead Time Forecasting for an INMA(1) Model

For the INMA(1) process of $Y_t = \beta \circ Z_{t-1} + Z_t$, the cumulative Y over lead time l is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1} = (\beta \circ Z_t + Z_{t+1}) + (\beta \circ Z_{t+1} + Z_{t+2}) + \cdots + (\beta \circ Z_{t+l} + Z_{t+l+1})$$

Equation 6-16

The above equation can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1} \sum_{i=1}^{n_j} \psi_{ij} \circ Z_{t+k_{ij}}$$

Equation 6-17

where n_j is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ and ψ_{ij} is the corresponding coefficient for each $Z_{t+k_{ij}}$.

It can be seen from Equation 6-16 that because the process is an integer moving average of order one, each of $\{Y_{t+j}\}_{j=1}^{l+1}$ only has two $Z_{t+k_{ij}}$ and therefore $n_j = 2$. The corresponding coefficient for each $Z_{t+k_{ij}}$, shown by ψ_{ij}, is $\{\psi_{1j} = \beta, \psi_{2j} = 1\}$. $t + k_{ij}$ is the subscript of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. From Equation 6-16 it can be seen that $\{k_{1j} = j - 1, k_{2j} = j\}$. Therefore,

$$k_{ij} = \begin{cases} j-1 & \text{for } i = 1 \\ j & \text{for } i = 2 \end{cases} \qquad \text{for } j = 1, \dots, l+1$$

Equation 6-18

All of these terms are shown in Table 6-3.

Based on Equation 6-17, the conditional expected value of the aggregated process is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \,|Y_t\right) = \left(\sum_{j=1}^{l+1} \sum_{i=1}^{2} \psi_{ij}\right)\lambda = \left(\sum_{j=1}^{l+1} (1+\beta)\right)\lambda = (l+1)(1+\beta)\lambda$$

Equation 6-19

Table 6-3 Coefficients of $Z_{t+k_{ij}}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INMA(1) model

$j=1$ $i=1,\dots,n_1$ where $n_1=2$	$\psi_{11}=1$ $\psi_{21}=\beta$	$k_{11}=1$ $k_{21}=0$
$j=2$ $i=1,\dots,n_2$ where $n_2=2$	$\psi_{12}=1$ $\psi_{22}=\beta$	$k_{12}=2$ $k_{22}=1$
$\vdots$	$\vdots$	$\vdots$
$j=l+1$ $i=1,\dots,n_{l+1}$ where $n_{l+1}=2$	$\psi_{1(l+1)}=1$ $\psi_{2(l+1)}=\beta$	$k_{1(l+1)}=l+1$ $k_{2(l+1)}=l$

6.3.3 Lead Time Forecasting for an INARMA(1,1) Model

In this section, the lead time forecast of an INARMA(1,1) process is derived. The results will be used in chapters 8 and 9. The aggregated process over lead time is:

$$\begin{aligned}\sum_{j=1}^{l+1} Y_{t+j} &= Y_{t+1}+Y_{t+2}+\cdots+Y_{t+l+1} = (\alpha\circ Y_t + Z_{t+1} + \beta\circ Z_t)\\ &+(\alpha\circ Y_{t-1} + Z_{t+2} + \beta\circ Z_{t+1}) + \cdots + (\alpha\circ Y_{t+l} + Z_{t+l+1} + \beta\circ Z_{t+l})\end{aligned}$$

Equation 6-20

The above equation can be written as:

$$\begin{aligned}\sum_{j=1}^{l+1} Y_{t+j} &= \left(\alpha\circ Y_t + \alpha^2\circ Y_t + \cdots + \alpha^{l+1}\circ Y_t\right) + \left(Z_{t+1} + \alpha\circ Z_{t+1} + \cdots + \alpha^l\circ Z_{t+1}\right)\\ &+\left(Z_{t+2} + \alpha\circ Z_{t+2} + \cdots + \alpha^{l-1}\circ Z_{t+2}\right) + \cdots + (Z_{t+l} + \alpha\circ Z_{t+l}) + Z_{t+l+1}\\ &+(\beta\circ Z_t + \alpha\beta\circ Z_t + \cdots + \alpha^l\beta\circ Z_t)\\ &+(\beta\circ Z_{t+1} + \alpha\beta\circ Z_{t+1} + \cdots + \alpha^{l-1}\beta\circ Z_{t+1}) + \cdots\\ &+(\beta\circ Z_{t+l-1} + \alpha\beta\circ Z_{t+l-1}) + \beta\circ Z_{t+l}\end{aligned}$$

Equation 6-21

The above result can be simplified to:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1} \alpha^j\circ Y_t + \sum_{j=1}^{l+1}\sum_{i=0}^{l+1-j} \alpha^j\circ Z_{t+j} + \sum_{j=1}^{l+1}\sum_{i=0}^{l+1-j} \alpha^j\beta\circ Z_{t+j-1}$$

Equation 6-22

Then, the conditional expected value of the Equation 6-21 is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\lambda(1+\beta)}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$

Equation 6-23

6.3.4 Lead Time Forecasting for an INARMA(*p*,*q*) Model

The results of lead time forecasting have been illustrated for INAR(1), INMA(1), INAR(2) and INARMA(1,2) processes (the last two are given in appendices 6.A and 6.B). This section investigates the lead time aggregation of an INARMA(*p*,*q*) process.

Proposition 1. Aggregation of an INARMA(*p*,*q*) process over a lead time results in an INARMA(*p*,*q*) process.

Proof.

For an INARMA(*p*,*q*) process of $Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i \circ Z_{t-i}$, the aggregated process over lead time can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1} \left\{\left[\sum_{i=1}^{p} \alpha_i \circ Y_{t+j-i}\right] + Z_{t+j} + \left[\sum_{i=1}^{q} \beta_i \circ Z_{t+j-i}\right]\right\} =$$
$$\sum_{i=1}^{p}\left[\alpha_i \circ \sum_{j=1}^{l+1} Y_{t+j-i}\right] + \sum_{j=1}^{l+1} Z_{t+j} + \sum_{i=1}^{q}\left[\beta_i \circ \sum_{j=1}^{l+1} Z_{t+j-i}\right]$$

Equation 6-24

Now, if we assume that $\sum_{j=1}^{l+1} Y_{t+j} = Y_\tau$ and $\sum_{j=1}^{l+1} Z_{t+j} = Z_\tau$, Equation 6-24 can be written as:

$$Y_\tau = \sum_{i=1}^{p} \alpha_i \circ Y_{\tau-i} + Z_\tau + \sum_{i=1}^{q} \beta_i \circ Z_{\tau-i}$$

Equation 6-25

which is also an INARMA(*p*,*q*) process. Therefore, aggregation of an INARMA(*p*,*q*) process over a lead time results in an INARMA(*p*,*q*) process with the same INAR and INMA parameters but with a different innovation parameter. Here, Z_τ is the sum of $(l+1)$ independent Poisson variables, thus $Z_\tau \sim Poi((l+1)\lambda)$.

Proposition 2. The over-lead-time-aggregated INARMA(p,q) process can be written in terms of the last p observations as follows:

$$\begin{aligned}\sum_{j=1}^{l+1} Y_{t+j} &= \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Y_{t-1} + \cdots \\ &+ \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^p} \psi_{ij}^p \circ Y_{t-p+1} + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^{p+1}} \psi_{ij}^{p+1} \circ Z_{t+k_{ij}}\end{aligned}$$

Equation 6-26

with the parameters as shown in Table 6-4 (see Appendix 6.C for the proof).

Table 6-4 Parameters of the over-lead-time-aggregated INARMA(p,q) model

for $w = 1, \dots, p$	$n_j^w = \begin{cases} \left(\sum_{i=1}^{p} n_{j-i}^w\right) + 1 & j \le p-(w-1) \\ \sum_{i=1}^{p} n_{j-i}^w & j > p-(w-1) \end{cases}$	$\psi_{ij}^w = \begin{cases} \begin{bmatrix} \alpha_p \psi_{i(j-p)}^w & i = 1, \dots, n_{j-p}^w \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^w & i = n_{j-2}^w + 1, \dots, n_{j-2}^w + n_{j-1}^w \\ \alpha_{j+(w-1)} & i = n_{j-1}^w + 1 \end{bmatrix} & j \le p-(w-1) \\ \begin{bmatrix} \alpha_p \psi_{i(j-p)}^w & i = 1, \dots, n_{j-p}^w \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^w & i = n_{j-2}^w + 1, \dots, n_{j-2}^w + n_{j-1}^w \end{bmatrix} & j > p-(w-1) \end{cases}$
	$n_j^{p+1} = \left(\sum_{i=1}^{p} n_{j-i}^{p+1}\right) + (q+1)$	$\psi_{ij}^{p+1} = \begin{cases} \alpha_p \psi_{i(j-p)}^{p+1} & i = 1, \dots, n_{j-p}^{p+1} \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^{p+1} & i = n_{j-2}^{p+1} + 1, \dots, n_{j-2}^{p+1} + n_{j-1}^{p+1} \\ \beta_q, \dots, \beta_1, 1 & i = n_{j-1}^{p+1} + 1, \dots, n_{j-1}^{p+1} + n_j^{p+1} \end{cases}$ $k_{ij} = \begin{cases} \{k_{i(j-p)}\} & i = 1, \dots, n_{j-p}^{p+1} \\ \vdots & \vdots \\ \{k_{i(j-1)}\} & i = \sum_{z=2}^{p} n_{j-z}^{p+1} + 1, \dots, \left(\sum_{z=2}^{p} n_{j-z}^{p+1}\right) + n_{j-1}^{p+1} \\ j-q, \dots, j-1, j & i = \sum_{z=1}^{p} n_{j-z}^{p+1} + 1, \dots, n_j^{p+1} \end{cases}$

Now, in order to find the forecast over lead time, we need to calculate the expected value of the aggregated process given the p-previous observations.

$$\begin{aligned}E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_{t-p+1}, \dots, Y_{t-1}, Y_t\right) &= \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2\right) Y_{t-1} + \cdots \\ &+ \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^p} \psi_{ij}^p\right) Y_{t-p+1} + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^{p+1}} \psi_{ij}^{p+1}\right) \lambda\end{aligned}$$

Equation 6-27

6.4 Conclusions

Forecasting with an INARMA process is discussed in this chapter. The minimum mean square error (MMSE) forecasts for INAR(*p*), INMA(*q*) and INARMA*(p,q)* processes are reviewed. This includes both one-step and *h*-step ahead forecasts. These forecasts are based on the conditional expected value of the process and, as argued by McCabe and Martin (2005), these are not coherent forecasts. This means that the results are not necessarily integers. However, using the conventional forecasting method of conditional expectations is the most widely used approach in the literature even for count series.

It is shown in this chapter that the aggregation of an INARMA*(p,q)* process with Poisson innovations (with mean λ) over a lead time l results in an INARMA*(p,q)* process with the same autoregressive and moving average parameters and the innovation parameter of $(l+1)\lambda$. The lead time aggregation and forecasting for the INARMA*(p,q)* process is obtained. In order to understand the implications of the results, some examples including a range of autoregressive and moving average processes are provided.

It will be discussed in chapter 7 that four INARMA models will be used in simulation and empirical analysis of this thesis. These models are INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1). Therefore, the lead time forecast for the last three processes are presented in this chapter. The lead time forecast of an INARMA(0,0) can simply be obtained from $(l+1)\lambda$.

Chapter 7 SIMULATION DESIGN

7.1 Introduction

This chapter addresses a simulation experiment based on theoretically generated data. A model-based simulation shows the evolution through time of a stochastic process, represented by a mathematical model through multiple realizations of the process. In this research, simulation is used for various reasons including:

- to assess the effects of the approximations made for the mathematical model
- to test the performance of identification methods
- to measure the accuracy of estimates of the model's parameters
- to assess the sensitivity of forecast accuracy to control parameters such as the number of observations and the sparsity of data (based on the INARMA parameters)

- to compare the forecasts of the mathematical model with other benchmark methods.

The chapter is organized as follows. The reasons for conducting simulation are discussed in section 7.2. The simulation design is defined in section 7.3, including the range of INARMA models to be used in the simulation, the control parameters and the performance metrics. Verification of the simulation is discussed in section 7.4 and, finally, section 7.5 provides the conclusions.

7.2 Rationale for Simulation

In chapter 4, two identification methods were discussed, namely the two-stage and the one-stage methods. Simulation enables us to find the percentage of theoretically generated INARMA time series that can be identified correctly by each of these methods. A further application of the simulation model is to investigate the effect of identifying an incorrect model for a specific series, or misidentification, on the accuracy of forecasts.

The next step in the INARMA methodology is estimating the parameters of the identified model. As explained in detail in chapter 5, Conditional Least Squares (CLS) and Yule-Walker (YW) are the two estimation methods used (CML will also be used for the INAR(1) process). The role of simulation is to compare the results of these methods in terms of: (*i*) how close are the estimates to the real parameters, which are known when theoretically generated data are being used, and (*ii*) which estimation method results in better forecasts.

The simulation model will be based on the assumption that the distribution of the innovations is Poisson. Although other distributions have been proposed in the literature including compound Poisson (McKenzie, 2003), negative Binomial (McKenzie, 1985; Al-Osh and Alzaid, 1987; Brännäs and Hall, 2001) and the Geometric (McKenzie, 1986; Alzaid and Al-Osh, 1988), this research only focuses on the Poisson. The sensitivity of the results to the distributional assumption can be analyzed but this will not be covered in this thesis. Other marginal distributions are beyond this research’s scope. The Poisson distribution is probably the most commonly used distribution in modelling counting processes (Alzaid and Al-Osh, 1990). It is the only distribution among the class of discrete self-decomposable[1] distributions which has a finite mean (Silva and Oliveira, 2004). Another property of interest is that in the INAR(1) and INMA(q) processes, the Poisson distribution plays a role similar to that of the Gaussian distribution in the AR(1) process. However, Jung and Tremayne (2006b) argued that only an INAR(2)-AA process with Poisson innovations results in a process with Poisson marginal distribution and the same is not true for an INAR(2)-DL process. Another advantage of the Poisson over other distributions is that it has only one parameter to estimate.

One of the main concerns in forecasting intermittent series is the length of available data history. This is because in practice we may be limited by short length of history. For example, the 3,000 series that we use in empirical analysis (see chapter 9) only has 24 periods of monthly data. Simulation enables us to check the sensitivity of the identification, estimation, and forecasting results to the length of the series.

Once the forecasting results have been established, simulation can be used to compare these results with benchmark methods, Croston, SBA, and SBJ methods (see chapter 2 for detailed discussion on benchmark methods). This includes one-step ahead, h-step ahead, and lead time forecasts.

In a nutshell, simulation is conducted to analyze the sensitivity of results to: the sparsity of data, the length of history, the parameters' ranges, the estimation methods, and the effect of misidentification. It also enables us to compare the INARMA forecasts with those of benchmark methods using different accuracy measures.

7.3 Simulation Design

7.3.1 The Range of Series

Different integer autoregressive moving average processes will be used to test the

[1] A distribution with probability generating function (p.g.f) P is called discrete self-decomposable if:
$P(z) = P(1 - \alpha + \alpha z)P_\alpha(z) \qquad |z| \leq 1 \qquad \alpha \in (0,1)$
where P_α is a p.g.f. The above equation can also be written in the form of:
$X = \alpha \circ X' + X_\alpha$
where $\alpha \circ X'$ and X_α are independent and X' is distributed as X (Sueutel, F. W. and K. van Harn (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**(5): 893-899.).

mathematical findings. We consider an INAR process, an INMA process and a mixed INARMA process. In order to test the performance of the benchmark methods, the special case of INARMA(0,0) (or simply an i.i.d. Poisson process) is also used. Therefore, the following four processes are assumed for this study:

INARMA(0,0), INARMA(1,0), INARMA(0,1), INARMA(1,1).

An extension to this study would be to examine higher order INARMA processes. However, as shown in a later chapter, the simpler models ((0,0) and (1,0)) perform very well on empirical data. Using the above models also has the benefit of having few parameters to be estimated.

7.3.2 Producing INARMA(*p*,*q*) Series

Since it has been assumed that the innovations are Poisson distributed ($Z_t \sim Poi(\lambda)$), we first need to generate i.i.d. Poisson random numbers.

The simulation code is written in MATLAB 6.1. Hence, we use the *poissrnd* function from MATLAB's statistics toolbox. The performance of this function is tested by the Poisson dispersion and the score tests (see section 4.2.2) and the results confirm the accuracy of the function.

Next, by assuming the values of autoregressive parameters $\{\alpha_j\}_{j=1}^{p}$ and moving average parameters $\{\beta_j\}_{j=1}^{q}$, the autoregressive and moving average components are generated using a Binomial random number generator. This is because, based on the properties of binomial thinning discussed in chapter 3, $\alpha \circ X$ given X has a binomial distribution with parameters (α, X). Therefore, for example in an INAR(1) model $(Y_t = \alpha \circ Y_{t-1} + Z_t)$, $\alpha \circ Y_{t-1}$ is obtained from generating a random Binomial number with parameters (Y_{t-1}, α). The Binomial numbers are generated using the *binornd* function from the MATLAB's statistics toolbox as a sum of Bernoulli random variables. The performance of this function is also tested by the score test and a built-in goodness-of-fit test (based on chi-square). The results, again, support the use of this function.

Then, the INARMA series is generated from the model:

$$Y_t = \sum_{j=1}^{p} \alpha_j \circ Y_{t-j} + Z_t + \sum_{j=1}^{q} \beta_j \circ Z_{t-j}$$

Equation 7-1

In order to obtain a stationary series, the series is initialized with the expected value of each process (Al-Osh and Alzaid, 1987; Brännäs, 1994). The expected value of the above process is given by:

$$E(Y_t) = \frac{\lambda\left(1 + \sum_{j=1}^{q} \beta_j\right)}{1 - \sum_{j=1}^{p} \alpha_j}$$

Equation 7-2

7.3.3 Control Parameters

The control parameters of the simulation are: the mean of the Poisson innovations (λ), autoregressive and moving average parameters ($\{\alpha_j\}_{j=1}^{p}, \{\beta_j\}_{j=1}^{q}$), the length of the series (n), the forecast horizon (h), the length of the lead time (l), and the benchmark methods' parameters. In this section, the ranges of these control parameters are reviewed.

7.3.3.1 INARMA Parameters

From the definition of the thinning operation it is obvious that the autoregressive and moving average parameters represent the chance of surviving for elements of the process at time $t-1$ (Y_{t-j} s and Z_{t-j} s, respectively). Therefore, these parameters are probabilities and can only take values in the range [0,1].

Other restrictions have to be applied on the autoregressive and moving average parameters in order to assure the stationarity and invertibility of the process. Table 7-1 reviews the range of values that these parameters can take for the INARMA processes selected in section 7.3.1 (see section 3.3.8 for stationarity and invertibility conditions of an INARMA(p,q) process).

Table 7-1 Range of autoregressive and moving average parameters

INARMA(p,q) models	**Range of autoregressive parameters**	**Range of moving average parameters**
INARMA(0,0)	-	-
INARMA(1,0)	$0 \leq \alpha < 1$ stationarity condition: $\alpha \neq 1$	-
INARMA(0,1)	-	$0 \leq \beta < 1$ invertibility condition: $\beta \neq 1$
INARMA(1,1)	$0 \leq \alpha < 1$ stationarity condition: $\alpha \neq 1$	$0 \leq \beta < 1$ invertibility condition: $\beta \neq 1$

The range of parameters for some simulation studies reported in the literature are reviewed in Table 7-2.

Table 7-2 Range of INARMA parameters studied in the literature

Study	**Number of observations $\boldsymbol{n}$**	**AR or MA parameter $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$**	**Innovation parameter $\boldsymbol{\lambda}$**	**Number of replications**
Al-Osh and Alzaid (1987)	$n = 50, 75, 100, 200$	$\alpha = 0.1, 0.2, 0.9$	$\lambda = 1\ (0.5), 3$	200
Brännäs and Hall (2001)	$n = 10\ (10), 100\ (100), 500$	$\beta_j = 0.1, 0.5, 0.9$	$\lambda = 5$	1000
Brännäs and Hellström (2001)	$n = 50\ , 100\ , 200$	$\alpha = 0.5, 0.7, 0.9$	$\lambda = 5, 10$	1000
Silva and Oliveira (2004)	$n = 64, 128, 512, 1024$	$\alpha = 0.1, 0.5, 0.9$	$\lambda = 1, 3$	200
Silva et al. (2005)	$n = 25, 50, 100$	$\alpha = 0.1, 0.3, 0.7, 0.9$	$\lambda = 1, 3$	
Bu et al. (2008)	$n = 100, 500$	$\alpha_j = 0.1, 0.3, 0.5, 0.7$	$\lambda = 1$	1000

Based on the constraints of Table 7-1, and taking into account previous experiments (Table 7-2), the parameter space for the four selected INARMA models used in this thesis is shown in Table 7-3.

If the discrete variates are large numbers, they can be approximated by continuous variates. It is when they are relatively small integers that using integer autoregressive moving average models becomes justifiable (McKenzie, 2003). Therefore, the innovation term (λ) has to be defined to assure the observations are small integers. As can be seen from Table 7-3, we assume a range of $\lambda = [0.5,5]$ for most models. For INARMA(0,0) we consider two other values of $\lambda = 0.3$ and $\lambda = 20$. This is to test the Croston-SBA categorization for highly intermittent and barely intermittent series.

Table 7-3 Parameter space for the selected INARMA models

INARMA(*p*,*q*) models	**Parameters**			
INARMA(0,0)	$\lambda = 0.3, 0.5, 0.7, 1, 3, 5, 20$			
INARMA(1,0)	$\alpha = 0.1, \lambda = 0.5$ $\alpha = 0.5, \lambda = 0.5$ $\alpha = 0.9, \lambda = 0.5$	$\alpha = 0.1, \lambda = 1$ $\alpha = 0.5, \lambda = 1$ $\alpha = 0.9, \lambda = 1$	$\alpha = 0.1, \lambda = 3$ $\alpha = 0.5, \lambda = 3$ $\alpha = 0.9, \lambda = 3$	$\alpha = 0.1, \lambda = 5$ $\alpha = 0.5, \lambda = 5$ $\alpha = 0.9, \lambda = 5$
INARMA(0,1)	$\beta = 0.1, \lambda = 0.5$ $\beta = 0.5, \lambda = 0.5$ $\beta = 0.9, \lambda = 0.5$	$\beta = 0.1, \lambda = 1$ $\beta = 0.5, \lambda = 1$ $\beta = 0.9, \lambda = 1$	$\beta = 0.1, \lambda = 3$ $\beta = 0.5, \lambda = 3$ $\beta = 0.9, \lambda = 3$	$\beta = 0.1, \lambda = 5$ $\beta = 0.5, \lambda = 5$ $\beta = 0.9, \lambda = 5$
INARMA(1,1)	$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$ $\alpha = 0.1, \beta = 0.9, \lambda = 0.5$ $\alpha = 0.5, \beta = 0.5, \lambda = 0.5$ $\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	$\alpha = 0.1, \beta = 0.1, \lambda = 1$ $\alpha = 0.1, \beta = 0.9, \lambda = 1$ $\alpha = 0.5, \beta = 0.5, \lambda = 1$ $\alpha = 0.9, \beta = 0.1, \lambda = 1$	$\alpha = 0.1, \beta = 0.1, \lambda = 5$ $\alpha = 0.1, \beta = 0.9, \lambda = 5$ $\alpha = 0.5, \beta = 0.5, \lambda = 5$ $\alpha = 0.9, \beta = 0.1, \lambda = 5$	

7.3.3.2 Length of Series

Different lengths of series are considered in order to test the sensitivity of results (identification, estimation, and forecasts accuracy) to the length of history. Because in real cases, we are often restricted by the short lengths of history (as will be seen in empirical analysis of this thesis) we use $n = 24, 36, 48, 96$. Only for investigating the accuracy of estimates in terms of bias and MSE (section 8.3), $n = 500$ is also added to the above cases.

The first half of the observations is assigned for identification and estimation, and is referred to as the estimation period. This also includes the benchmark methods of Croston, SBA and SBJ. The second half is left for forecasting and is called the performance period.

7.3.3.3 Forecast Horizon and Lead Time

Three-step and six-step ahead forecasts are calculated in addition to one-step ahead forecasts. The lead times considered are also three and six periods.

The number of replications is set to 1000. However, for the INARMA(0,0) model with very small mean ($\lambda \leq 1$) more replications are used to reduce the sampling error. Therefore, the number of replications for $\lambda = 0.3, 0.5$ is 30,000 and for $\lambda = 0.7, 1$ is 10,000.

7.3.3.4 Benchmark Methods' Parameters

As discussed in chapter 2, three methods of forecasting intermittent demand are selected to compete against the INARMA method. These methods are: Croston (Croston, 1972), SBA (Syntetos and Boylan, 2005) and SBJ (Shale et al., 2006).

All of these methods are based on separate smoothing of demand sizes and on the interval between positive demands using a common smoothing parameter for size and interval. Therefore a smoothing constant needs to be selected. It has been suggested in the literature (e.g. Brown, 1959; Croston, 1972) that, especially when the length of history is short, it is best to use fixed values of the smoothing parameter.

We choose two arbitrary values for smoothing parameter: $\alpha = 0.2$ and $\alpha = 0.5$. The first value is selected because in intermittent demand context low smoothing constant values are suggested (Syntetos and Boylan, 2005). However, as can be seen from Table 7-3, some generated series have high autocorrelation; therefore $\alpha = 0.5$ is also used.

For initialization of the methods, the first inter-demand interval is used as the first smoothed inter-demand interval. For the first smoothed size, the average of the first two positive demands is used. If fewer than two positive demands is observed in the estimation period, the estimation period for that particular replication is extended until two non-zero demands are observed.

7.3.4 Identification Procedure

As argued in chapter 4, the sample autocorrelation function (SACF) and partial autocorrelation function (SPACF) of INARMA models have the same structure as those of ARMA models and therefore can be used in identifying the moving average and autoregressive orders of the model. However, as argued in section 4.6, for simulation purposes automated methods such as penalty functions should be used.

Jung and Tremayne (2003) argue that the first step in analysing time series of counts is to investigate if the data exhibit any serial dependence. If such dependence does not exist, standard methods for independent data should be used. Based on this argument, two identification procedures were suggested in chapter 4, namely, two-stage and one-stage methods.

In the two-stage identification method, a Ljung-Box test of Equation 4-6 is first used to test if data has serial dependence. The reasons for the selection of this test were discussed in section 4.6. The second step involves using the AIC of Equation 4-28 (or where applicable, AIC_C of Equation 4-29) to select the appropriate model among the three possible INARMA models (see section 4.6 for discussion on the application of AIC of ARMA models for INARMA series).

In the one-stage identification method, the first step of the previous method is ignored. This means that the AIC is used to select among all possible INARMA models (INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1)).

The results of these two methods will be compared in terms of the percentage of series for which the model is identified correctly. This can be done in simulation because the correct model from which the series is produced is known. Another aspect that can be tested is the accuracy of forecasts obtained from each identification method (the accuracy measures are reviewed in section 7.3.7).

No identification method can guarantee that the correct model is identified at all times. In such cases, the effect of misidentification on the accuracy of forecasts is of interest. This will also be tested in the next chapter.

7.3.5 Estimation of Parameters

As discussed in chapter 5, Yule-Walker (YW) and conditional least squares (CLS) methods are used for estimation of parameters of INAR(1), INMA(1), and INARMA(1,1) processes. Because the conditional maximum likelihood (CML) estimation has been established only for INAR(p) processes, we can use it only for an INAR(1) process. For an INARMA(0,0) process, the three estimation methods result in the same estimator. All of these estimators are given in chapter 5.

The performance of these estimators has been tested in the literature (Al-Osh and Alzaid, 1987; Brännäs, 1994; Bu, 2006). However, this has been done for sample sizes greater than 50. Because we also use smaller numbers of observations, we compare the performance of these estimators. Since the true values of the parameters are known in simulation, we compare the bias and the MSE of the estimates. The impact of estimates on forecast accuracy is also an important issue that has not been looked at before and is covered in this thesis.

7.3.6 Forecasting Method

This thesis focuses on comparing the accuracy of forecasts produced by INARMA and benchmark methods. The accuracy measures include MSE and MASE (see section 7.3.7). We use the conditional expected value which yields minimum mean square error (MMSE) forecasts. It has been argued in the literature that this method is not coherent in that it does not produce integer-valued forecasts (Freeland and McCabe, 2004b). Other methods such as conditional median, Markov Chains, and bootstrapping have been suggested to tackle this problem (Cardinal et al., 1999; Freeland and McCabe, 2004b; Jung and Tremayne, 2006b; Bu and McCabe, 2008). However, none of these methods produces MMSE forecasts. Also, those methods that produce the distribution forecast instead of point forecasts are not used for our comparison. Such methods are definitely useful for competing against bootstrap methods for intermittent demand forecasting such as Willemain's bootstrap (Willemain et al., 2004) and can be considered as a future line of study.

The h-step ahead forecasts and lead time forecasts for INARMA models are discussed in chapter 6 in detail. The Croston, SBA and SBJ forecasts are given in chapter 2. For these methods, the h-step ahead forecasts are the same as the one-step ahead forecasts and the lead time forecast is simply the one-step ahead forecast multiplied by the length of lead time.

Finally, two cases regarding the forecast timing are considered: all points in time or focusing on those periods immediately after a positive demand is occurred (issue points). This is because Croston's method is designed to outperform the SES for issue points and it is of interest to test the performance of the INARMA method for issue points.

7.3.7 Performance Metrics

In this section, the performance measures to be used in the simulation are reviewed. In the identification stage, where we want to examine the capability of the two identification procedures, the percentage of correctly identified models is calculated. The accuracy of the forecasts produced by each identification method is also compared.

In order to compare the estimation methods (YW, CLS and CML only for INAR(1)), the bias (using Mean Error) and Mean Square Error (MSE) of parameters' estimates are calculated. The performance of the estimates is also compared in terms of their impact on forecast accuracy.

Finally, selecting the appropriate forecasting accuracy measure is an important issue for intermittent processes. As discussed in section 2.4, the fact that intermittent demand series include zeros, makes some of the conventional measures inappropriate. The following accuracy measures are used in this thesis: Mean Error (ME), Mean Square Error (MSE), Mean Absolute Scaled Error (MASE) for simulation, along with Percentage Better (PB) of MASE and Relative Geometric Root-Mean-Square Error (RGRMSE) for empirical analysis (see section 2.4.3 for more details).

7.4 Verification

Verification is the process to make sure that no programming error has been made (Kleijnen and Groenendaal, 1992). This can be done by calculating some intermediate results manually and comparing them with the results obtained by the program. This is called *tracing* (Kleijnen and Groenendaal, 1992). *Eyeballing* or reading through the code and looking for *bugs* is another way of verification (Kleijnen and Groenendaal, 1992). The following steps have been done in order to verify the simulation model:

- The MATLAB code has been read through to make sure that the correct logic and functions have been used.
- The intermediate and also the final results have been compared for a limited number of replications (e.g. 20 replications) with MS Excel.
- The average and standard deviation of the generated INARMA series is calculated and compared to the theoretical mean and standard deviation of the process to test the generated data.

The selection of parameters was made to make sure that both highly-intermittent and less-intermittent data are considered. Inter-arrival times are also obtained for each time series.

7.5 Conclusions

In this chapter, a simulation experiment was developed to assess the accuracy of approximations made for the mathematical analysis, to measure the accuracy of estimates, to assess the sensitivity of forecast accuracy to control parameters, and to compare the INARMA forecasts with those of benchmark methods.

Four integer autoregressive moving average models have been selected for the purpose of simulation (models with $p, q \leq 1$). The marginal distribution is assumed to be Poisson. The control parameters used are: autoregressive and moving average parameters, innovation parameter, forecast horizon, the length of lead time, and the smoothing parameter for the benchmark methods.

As previously discussed in section 2.4.3, different accuracy measures are needed to assess the accuracy of estimates and forecasts. The accuracy of estimates is measured using ME and MSE. Demand being intermittent makes some forecast accuracy measures not applicable. We have selected ME, MSE, and MASE for simulation. The PB of MASE and RGRMSE will be added to the above measures for empirical analysis.

Chapter 8 SIMULATION RESULTS

8.1 Introduction

The simulation results are presented in this chapter. As discussed in chapter 7, the main objective of simulation is to test whether using an INARMA model results in better forecasts compared to benchmark methods of Croston, Syntetos-Boylan Approximation (SBA) and Shale-Boylan-Johnston (SBJ). This is discussed in section 8.6. The simulation experiment also enables us to test the applicability of the Croston-SBA categorization (Syntetos et al., 2005) when demand is an INARMA process.

As discussed in section 7.3.1, four processes are simulated: INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1). Based on the arguments in section 2.4, the ME, MSE and MASE of the forecasts are compared to those of Croston, SBA and

SBJ methods. A range of INARMA parameters and different lengths of history are used (see section 7.3.3).

The estimation methods used in this study are YW and CLS for INAR(1), INMA(1) and INARMA(1,1) processes and CML for INAR(1) (see Chapter 5 for detailed discussion). As another objective of simulation, the accuracy of parameters' estimates needs to be tested. The performance of the estimators can be tested not only by comparing the accuracy of the estimates, but also by comparing their impact on the forecast accuracy. The former has been undertaken by comparing the ME and MSE of the parameters' estimates (see section 8.3 and Appendix 8.A). The latter, the results of which are presented in section 8.4, has been accomplished by comparing the ME, MSE and MASE of forecasts obtained using each estimation method.

The chapter is structured as follows. Details of the simulation design are reviewed in section 8.2. Sections 8.3 and 8.4 compare the accuracy of different estimates of the parameters of INARMA processes. The Croston-SBA categorization (Syntetos et al., 2005) for data produced by INARMA models is validated in section 8.5. The INARMA forecasts are then compared to the benchmark methods in section 8.6. It is first assumed that the order of the INARMA model is known. The results for the case where the order needs to be identified are presented in section 8.6.2. The lead-time forecasts are compared in section 8.6.3 and the conclusions are provided in section 8.7.

8.2 Details of Simulation

As mentioned in chapter 7, the number of replications is set to 1000. However, in order to reduce the sampling error for the case of INARMA(0,0) process with small parameters, ($\lambda = 0.3, 0.5$) and ($\lambda = 0.7, 1$), 30,000 and 10,000 replications are used, respectively.

It has been suggested in the literature that, especially with short length of history, it is best to use fixed values of smoothing parameters (Brown, 1959; Croston, 1972). Because with intermittent demand, data history is short in most cases, we use two arbitrary values for the smoothing parameter for Croston, SBA and SBJ ($\alpha = 0.2$ and

$\alpha = 0.5$).

As summarized in chapter 7, the initialization for Croston, SBA and SBJ is based on using the first inter-demand interval as the first smoothed inter-demand interval and the average of the first two non-zero observations as the first smoothed size. The observations are divided into two categories: estimation period and performance period. Initialization and estimation of parameters are conducted in the estimation period and the estimates' accuracy and forecasting accuracy are assessed in the performance period. If at least two non-zero demands are observed in the estimation period, the first half of the observations is assigned for the estimation period and the other half for the performance period. However, if fewer than two non-zero demands are observed in the estimation period, this period will be extended until the second non-zero demand is observed.

In order to obtain a stationary series, we initialize the INARMA methods with the expected value of each model. As discussed in chapter 7, the forecasting accuracy is obtained for both cases of all points in time and issue points (i.e. after a positive demand is observed). Finally, if there is no nonzero observation in the performance period, the error measures for issue points are excluded (only for the corresponding replication). If the in-sample MAE is zero, the MASE for that replication is excluded.

8.3 Accuracy of INARMA Parameter Estimates

As previously discussed in chapter 5, two methods (YW and CLS) have been used to estimate the parameters of all four INARMA processes. In this section, the accuracy of these parameter estimates is evaluated using MSE. Out of the four INARMA processes of this study, only three are included for comparison of estimation methods: INAR(1), INMA(1) and INARMA(1,1). The YW, CLS and CML estimates for INARMA(0,0) are the same (see section 5.2).

As previously mentioned, CML is also used in addition to YW and CLS in order to estimate the parameters of an INAR(1) process. The reason for excluding CML for other processes is that the maximum likelihood functions for INMA(1) and

INARMA(1,1) processes have not been developed in the literature (see Chapter 5).

The parameters may fall out of the region $[0,1]$. In order to tackle this issue, the parameters are set equal to their closest boundary value in each case (Brännäs and Hall, 2001).

The accuracy of YW, CLS and CML estimates of the parameters of an INAR(1) process for the case of $n = 24$ are compared in Table 8-1. For high values of n and also when the mean of the process, $\lambda/(1-\alpha)$, is high, the CML becomes computationally expensive.

Table 8-1 MSE of YW, CLS and CML estimates for INAR(1) series when $n = 24$

Parameters	α			λ		
	YW	**CLS**	**CML**	**YW**	**CLS**	**CML**
$\alpha = 0.1, \lambda = 0.5$	0.0203	0.0238	0.0337	0.0391	0.0394	0.0447
$\alpha = 0.5, \lambda = 0.5$	0.0772	0.0762	0.0734	0.1381	0.1337	0.1043
$\alpha = 0.9, \lambda = 0.5$	0.1436	0.1056	0.0042	3.9057	2.8683	0.0891
$\alpha = 0.1, \lambda = 1$	0.0196	0.0226	0.0364	0.0877	0.0910	0.1104
$\alpha = 0.5, \lambda = 1$	0.0723	0.0704	0.0631	0.4136	0.3950	0.3125
$\alpha = 0.9, \lambda = 1$	0.1429	0.1067	0.0028	14.7487	11.0440	0.2526
$\alpha = 0.1, \lambda = 3$	0.0188	0.0215	0.0419	0.4225	0.4509	0.6832
$\alpha = 0.5, \lambda = 3$	0.0716	0.0684	0.0658	2.9762	2.8352	2.6277
$\alpha = 0.9, \lambda = 3$	0.1462	0.1124	0.0024	134.0715	102.9696	2.0292
$\alpha = 0.1, \lambda = 5$	0.0197	0.0227	0.0390	0.9650	1.0462	1.6005
$\alpha = 0.5, \lambda = 5$	0.0710	0.0686	0.0606	7.6814	7.4755	6.4135

The results confirm that, as suggested by Al-Osh and Alzaid (1987), the MSE of estimates produced by CML is generally less than that of YW and CLS (with the exception of the cases where $\alpha = 0.1$). However, it will be seen in a later section that the results of CML in terms of its effect on forecast accuracy are not very far from those by YW and CLS. This is also true for those cases in Table 8-1 that the MSE of CML is much less than that of the other methods (e.g. $\alpha = 0.9$ and $\lambda = 3$).

The results of comparing the MSE of YW and CLS estimates of the parameters of INAR(1), INMA(1) and INARMA(1,1) processes are shown in Table 8-2, Table 8-3, and Table 8-4, respectively.

Al-Osh and Alzaid (1987) suggest that the accuracy of YW and CLS estimates for parameters of an INAR(1) process are close. The results of Table 8-2 confirm this when the number of observations is high. However, for fewer observations, the difference is high when the autoregressive parameter is high.

Table 8-2 Accuracy of YW and CLS estimates for INAR(1) series

Parameters	$\text{MSE}(\hat{\alpha}_{YW})/\text{MSE}(\hat{\alpha}_{CLS})$					$\text{MSE}(\hat{\lambda}_{YW})/\text{MSE}(\hat{\lambda}_{CLS})$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.8402	0.8978	0.9351	0.9700	1.0000	0.9868	0.9918	0.9890	1.0000	1.0000
$\alpha = 0.5, \lambda = 0.5$	1.0291	1.0291	1.0199	1.0248	1.0000	1.0408	1.0289	1.0194	1.0182	1.0000
$\alpha = 0.9, \lambda = 0.5$	1.3036	1.3487	1.3063	1.2404	1.1250	1.3089	1.3355	1.2935	1.2440	1.0874
$\alpha = 0.1, \lambda = 1$	0.8522	0.9222	0.9388	0.9691	1.0000	0.9655	0.9728	0.9762	0.9891	1.0000
$\alpha = 0.5, \lambda = 1$	1.0255	1.0227	1.0278	1.0286	1.0000	1.0213	1.0348	1.0227	1.0230	1.0000
$\alpha = 0.9, \lambda = 1$	1.3502	1.2990	1.3029	1.2294	1.1429	1.3550	1.3101	1.3004	1.2291	1.0957
$\alpha = 0.1, \lambda = 3$	0.8610	0.9102	0.9379	0.9655	1.0000	0.9228	0.9422	0.9644	0.9806	0.9946
$\alpha = 0.5, \lambda = 3$	1.0319	1.0349	1.0224	1.0148	1.0000	1.0317	1.0336	1.0307	1.0185	1.0024
$\alpha = 0.9, \lambda = 3$	1.3285	1.3063	1.3038	1.2526	1.1429	1.3273	1.3012	1.3010	1.2485	1.0954
$\alpha = 0.1, \lambda = 5$	0.8649	0.9118	0.9416	0.9670	1.0000	0.9094	0.9370	0.9565	0.9776	0.9959
$\alpha = 0.5, \lambda = 5$	1.0290	1.0396	1.0224	1.0216	1.0000	1.0355	1.0400	1.0240	1.0204	1.0068
$\alpha = 0.9, \lambda = 5$	1.3288	1.2945	1.3123	1.2376	1.1429	1.3236	1.2904	1.3110	1.2391	1.0960

The results show that for an INAR(1) process, when the number of observations is small, for high values of α, CLS produces much better estimates for both α and λ in terms of MSE (up to 35 percent improvement in MSE). On the other hand, for small values of α, YW results in better estimates (up to 16 percent improvement in MSE). The results of section 8.4 show that this is also true for the accuracy of forecasts produced by these estimates.

The MSE of $\hat{\lambda}$ for both YW and CLS estimates increases with an increase in α but this is not necessarily the case for the MSE of $\hat{\alpha}$ (see Appendix 8.A). This confirms the argument by Al-Osh and Alzaid (1987).

The results of Table 8-3 show that for an INMA(1) series, for a small number of observations, CLS has smaller MSE than YW except for the case of $\beta = 0.9$. When the number of observations increases, for high values of β, the MSE of YW estimates decreases with a greater pace compared to CLS. However, as will be discussed in section 8.4, it does not have a great effect on the accuracy of forecasts produced by each method. The MSE of $\hat{\lambda}$ for both YW and CLS estimates increases with an increase in β but the same is not necessarily true for the MSE of $\hat{\beta}$ (see Appendix 8.A).

Table 8-3 Accuracy of YW and CLS estimates for INMA(1) series

Parameters	$\text{MSE}(\hat{\beta}_{YW})/\text{MSE}(\hat{\beta}_{CLS})$					$\text{MSE}(\hat{\lambda}_{YW})/\text{MSE}(\hat{\lambda}_{CLS})$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	2.0558	1.9491	2.0409	1.9897	1.3929	0.9838	1.0084	1.0265	1.0510	1.0500
$\beta = 0.5, \lambda = 0.5$	1.1761	1.2907	1.1692	1.1307	0.5213	0.9008	0.9189	0.9032	0.8563	0.6098
$\beta = 0.9, \lambda = 0.5$	0.8054	0.6843	0.6217	0.5340	0.2336	0.9462	0.8521	0.8112	0.7927	0.5000
$\beta = 0.1, \lambda = 1$	1.9449	2.0335	2.0395	1.7292	1.3793	0.9952	1.0249	1.0518	1.0909	1.0962
$\beta = 0.5, \lambda = 1$	1.1879	1.1665	1.2214	1.1247	0.5354	0.8940	0.9045	0.8977	0.8772	0.5968
$\beta = 0.9, \lambda = 1$	0.7304	0.6812	0.5904	0.4636	0.2263	0.8650	0.8296	0.8080	0.7352	0.4012
$\beta = 0.1, \lambda = 3$	2.2829	2.1860	2.1367	1.8646	1.4615	1.0799	1.1157	1.1649	1.2067	1.1815
$\beta = 0.5, \lambda = 3$	1.1216	1.1679	1.1264	1.0872	0.4928	0.8906	0.9078	0.9011	0.8400	0.5050
$\beta = 0.9, \lambda = 3$	0.6900	0.6048	0.5664	0.3868	0.1909	0.8005	0.7554	0.7359	0.6084	0.3300
$\beta = 0.1, \lambda = 5$	2.4159	2.4242	2.2715	1.8780	1.4074	1.1174	1.2171	1.2425	1.2495	1.1940
$\beta = 0.5, \lambda = 5$	1.0589	1.0776	1.1294	1.0726	0.4634	0.8654	0.8406	0.8873	0.8217	0.4519
$\beta = 0.9, \lambda = 5$	0.6060	0.5175	0.4474	0.3523	0.1727	0.7006	0.6274	0.5856	0.5120	0.2597

The results of Table 8-4 show that, for INARMA(1,1) series, CLS produces better estimates especially when the number of observations is small and the autoregressive parameter is high. This is also true for the accuracy of forecasts produced by CLS compared to those by YW (as shown later).

To conclude, for INAR(1), INMA(1), and INARMA(1,1) processes, the autoregressive and moving average parameters and the number of observations determine which estimation method produces more accurate estimates. For an INAR(1) process, CLS outperforms YW for high values of α. The same is generally true for an INMA(1) process with low values of β and small number of observations. Finally, for an INARMA(1,1) process, CLS generally produces better estimates than YW with a few exceptions.

8.4 Forecasting Accuracy of INARMA Estimation Methods

As previously discussed in chapter 5, two methods (CLS and YW) have been used to estimate the parameters of all four INARMA processes. In this section, the accuracy of these estimates in terms of their effect on forecast accuracy is evaluated. The forecast accuracy is measured by ME, MSE, and MASE (see section 2.4 for detailed discussion). We focus on MSE in this section. MSE is specially selected because of its theoretical tractability. Also due to the fact that data is theoretically generated, the scale-dependency problem is not an issue when we average across multiple series.

Table 8-4 Accuracy of YW and CLS estimates for INARMA(1,1) series

Parameters	$\mathrm{MSE}(\hat{\alpha}_{YW})/\mathrm{MSE}(\hat{\alpha}_{CLS})$					$\mathrm{MSE}(\hat{\beta}_{YW})/\mathrm{MSE}(\hat{\beta}_{CLS})$					$\mathrm{MSE}(\hat{\lambda}_{YW})/\mathrm{MSE}(\hat{\lambda}_{CLS})$				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	4.8614	5.4738	6.0749	7.0578	3.4853	1.7669	1.9854	1.9831	1.9153	1.6269	2.0720	2.6889	3.1239	3.9265	3.7576
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.8613	0.7344	0.6736	0.6839	0.9770	0.6373	0.5349	0.4602	0.3182	0.1052	1.0420	0.9860	0.9155	0.7370	0.3245
$\alpha=0.5,\beta=0.5,\lambda=0.5$	1.3778	1.5195	1.5427	1.4870	1.1607	1.2698	1.2589	1.3484	1.3281	0.5957	1.0321	0.9228	0.8333	0.6433	0.2514
$\alpha=0.9,\beta=0.1,\lambda=0.5$	2.1067	2.0306	1.8965	1.5271	1.2222	26.3171	27.4400	28.9136	20.4444	3.8571	1.2033	1.1088	1.0270	0.9594	0.8696
$\alpha=0.1,\beta=0.1,\lambda=1$	4.9796	6.3395	6.5819	5.9756	3.6176	2.0247	1.8398	1.7630	1.7983	1.5588	2.7496	3.5359	3.9433	4.7348	4.6538
$\alpha=0.1,\beta=0.9,\lambda=1$	0.8243	0.5925	0.6154	0.6291	0.8710	0.5975	0.4733	0.4273	0.2969	0.1013	1.0431	0.9933	0.9032	0.7263	0.2915
$\alpha=0.5,\beta=0.5,\lambda=1$	1.6988	1.7842	1.7343	1.6031	1.2157	1.0136	1.0992	1.1177	1.0436	0.5140	1.0337	0.9391	0.8374	0.6111	0.2446
$\alpha=0.9,\beta=0.1,\lambda=1$	2.3724	2.2172	2.0000	1.7658	1.3750	43.0375	35.2133	31.3239	19.9841	2.8772	1.1029	1.0416	0.9534	0.9135	0.8853
$\alpha=0.1,\beta=0.1,\lambda=5$	5.2880	6.5479	7.3636	8.0300	3.3701	2.5138	2.3421	2.1606	1.9722	1.6857	4.4277	5.6560	6.6359	8.6472	5.6930
$\alpha=0.1,\beta=0.9,\lambda=5$	0.5406	0.4332	0.3695	0.3471	0.4227	0.4187	0.3415	0.2803	0.1872	0.0582	1.1328	1.0459	0.9843	0.7719	0.2663
$\alpha=0.5,\beta=0.5,\lambda=5$	2.2217	2.3401	2.5780	2.1421	1.0517	0.8303	0.8293	0.8479	0.7495	0.4093	0.9961	0.9877	0.9281	0.7039	0.2679
$\alpha=0.9,\beta=0.1,\lambda=5$	2.5276	2.3845	2.2099	1.8218	1.5000	45.5789	37.9867	31.5135	17.3108	2.4211	1.0838	0.9427	0.8756	0.8296	0.8635

The effect of CLS and YW estimates on one-step ahead forecasts is presented in Table 8-5, Table 8-6, and Table 8-7. The results for three-step ahead and six-step ahead forecasts are presented in Table 8-8 to Table 8-13.

For an INAR(1) process, the results of YW and CLS are compared to those of CML for $n = 24$. For longer length of history, the CML results become computationaly expensive. The results for an INAR(1) process show that when the history is short and data is highly autocorrelated ($\alpha = 0.9$), CLS produces more accurate forecasts (up to 11 percent improvement in MSE) than YW. For $\alpha \leq 0.5$, YW produces better forecasts, but the magnitude of improvement is small (up to a maximum of 3 percent improvement in MSE). The results also confirm that when the number of observations increases, the two methods yield very similar forecast errors (Al-Osh and Alzaid, 1987; Bu, 2006).

Table 8-5 One-step ahead forecast error comparison (YW, CLS and CML) for INAR(1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$					$\mathrm{MSE_{YW}}/\mathrm{MSE_{CML}}$
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$
$\alpha = 0.1, \lambda = 0.5$	0.9752	0.9841	0.9936	0.9978	1.0000	0.9727
$\alpha = 0.5, \lambda = 0.5$	0.9847	0.9954	0.9974	0.9999	1.0001	0.9799
$\alpha = 0.9, \lambda = 0.5$	1.1087	1.0957	1.0560	1.0248	1.0010	1.2084
$\alpha = 0.1, \lambda = 1$	0.9774	0.9926	0.9956	0.9985	0.9999	0.9751
$\alpha = 0.5, \lambda = 1$	0.9871	1.0006	0.9984	0.9995	1.0001	0.9899
$\alpha = 0.9, \lambda = 1$	1.1268	1.0836	1.0637	1.0249	1.0012	1.2329
$\alpha = 0.1, \lambda = 3$	0.9842	0.9935	0.9962	0.9987	0.9999	0.9694
$\alpha = 0.5, \lambda = 3$	0.9877	1.0009	1.0019	1.0000	1.0001	0.9850
$\alpha = 0.9, \lambda = 3$	1.0993	1.0807	1.0568	1.0228	1.0009	1.2605
$\alpha = 0.1, \lambda = 5$	0.9849	0.9922	0.9960	0.9987	0.9999	0.9687
$\alpha = 0.5, \lambda = 5$	0.9925	1.0000	1.0004	1.0003	1.0001	1.0019
$\alpha = 0.9, \lambda = 5$	1.1133	1.0866	1.0647	1.0252	1.0011	-

As noted previously, it is computationally expensive to calculate CML when the mean of the process is high. Therefore, no result is presented for the last case in Table 8-5.

The results also show that, except for the case where the autoregressive parameter is high, YW forecasts have smaller MSE than CML forecasts for $n = 24$. The above discussion about YW and CLS suggests that for such cases CLS is better than YW, but the results show that CML is still better than CLS for these cases.

The results for an INMA(1) process show that the forecast accuracy of YW and CLS estimates are generally close. When the history is short, CLS produces better forecasts for lower values of λ (up to 1.4 percent improvement in MSE). As shown in Table 8-6, for high values of λ, YW outperforms CLS (up to 3 percent improvement in MSE).

Table 8-6 One-step ahead forecast error comparison (YW and CLS) for INMA(1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	1.0000	1.0003	1.0008	1.0004	1.0000
$\beta = 0.5, \lambda = 0.5$	1.0065	1.0080	1.0054	1.0034	1.0008
$\beta = 0.9, \lambda = 0.5$	1.0137	1.0156	1.0111	1.0069	1.0013
$\beta = 0.1, \lambda = 1$	0.9998	1.0009	1.0005	1.0004	1.0001
$\beta = 0.5, \lambda = 1$	1.0058	1.0040	1.0045	1.0037	1.0031
$\beta = 0.9, \lambda = 1$	1.0084	1.0062	1.0075	1.0019	1.0003
$\beta = 0.1, \lambda = 3$	0.9985	0.9999	1.0003	1.0004	1.0000
$\beta = 0.5, \lambda = 3$	1.0020	1.0032	1.0006	1.0019	0.9998
$\beta = 0.9, \lambda = 3$	0.9930	0.9888	0.9915	0.9914	0.9944
$\beta = 0.1, \lambda = 5$	0.9954	0.9986	0.9992	0.9999	0.9999
$\beta = 0.5, \lambda = 5$	0.9972	0.9957	0.9964	0.9990	0.9989
$\beta = 0.9, \lambda = 5$	0.9888	0.9850	0.9857	0.9835	0.9885

For an INARMA(1,1) process the results show that CLS always produces better forecasts than YW. As shown in Table 8-7, when $\alpha \leq 0.5$, CLS outperforms YW by up to 20 percent. The difference is much greater when $\alpha = 0.9$ (up to 90 percent improvement in MSE). However, with an increase in the number of observations, the two methods become closer, especially for $\alpha = 0.9$.

Table 8-7 One-step ahead forecast error comparison (YW and CLS) for INARMA(1,1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	1.0850	1.1076	1.1175	1.0991	1.0259
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.0273	1.0556	1.0551	1.0499	1.0185
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	1.0633	1.0842	1.0805	1.0467	1.0087
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.4006	1.2431	1.1654	1.0475	1.0032
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0992	1.1077	1.1015	1.0874	1.0252
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0527	1.0624	1.0700	1.0568	1.0188
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	1.1293	1.1172	1.1031	1.0565	1.0104
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.4797	1.2853	1.1771	1.0561	1.0031
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.1110	1.1142	1.1175	1.1074	1.0246
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	1.1254	1.1285	1.1397	1.1273	1.0693
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	1.1939	1.1694	1.1603	1.0789	1.0232
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.9098	1.4115	1.2442	1.0676	1.0042

Based on the above results for one-step ahead forecasts, for INMA(1), YW and CLS are close. For INAR(1), when the history is short and the autoregressive parameter is high, CLS is considerably better than YW. But when $\alpha \leq 0.5$, the difference is much smaller. For INARMA(1,1) and especially for short history, CLS estimates produce better results than YW.

Although the above results are based on MSE, using MASE produces similar results (see Appendix 8.B). For INAR(1), CLS produces more accurate forecasts (up to 9 percent improvement in MASE) when the history is short and the autoregressive parameter is high ($\alpha = 0.9$). For $\alpha \leq 0.5$, YW produces better forecasts, but the magnitude of improvement is small (up to a maximum of 3 percent improvement in MASE). For an INMA(1) process, the forecasting accuracy of YW and CLS forecasts using MASE are very close. For an INARMA(1,1) process, CLS produces better forecasts than YW in most of the cases (up to 30 percent improvement in MASE). Finally, the results confirm that when the number of observations increases, the two methods become very close in terms of MASE.

The results for three-step and six-step ahead forecasts for INAR(1) process are shown in Table 8-8 and Table 8-9. Although the three-step ahead results follow the same pattern as one-step ahead forecasts, both three-step and six-step ahead forecasts produced by YW and CLS estimates are very close.

The results of YW and CLS three-step and six-step ahead forecasts for INMA(1) process are presented in Table 8-10 and Table 8-11. It can be seen that the two estimation methods result in very close forecasts.

Table 8-8 Three-step ahead forecast error comparison (YW and CLS) for INAR(1) series

Parameters	MSE_{YW}/MSE_{CLS}			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \lambda = 0.5$	0.9988	0.9992	0.9997	0.9998
$\alpha = 0.5, \lambda = 0.5$	0.9906	0.9993	1.0009	1.0013
$\alpha = 0.9, \lambda = 0.5$	1.0081	1.0189	1.0229	1.0382
$\alpha = 0.1, \lambda = 1$	0.9989	0.9996	0.9996	1.0000
$\alpha = 0.5, \lambda = 1$	0.9965	0.9995	1.0001	1.0012
$\alpha = 0.9, \lambda = 1$	1.0181	1.0210	1.0170	1.0367
$\alpha = 0.1, \lambda = 3$	0.9982	0.9994	0.9998	0.9999
$\alpha = 0.5, \lambda = 3$	0.9933	0.9995	1.0003	1.0014
$\alpha = 0.9, \lambda = 3$	1.0070	1.0135	1.0245	1.0362
$\alpha = 0.1, \lambda = 5$	0.9990	0.9998	0.9999	0.9999
$\alpha = 0.5, \lambda = 5$	0.9938	0.9997	1.0005	1.0014
$\alpha = 0.9, \lambda = 5$	1.0033	1.0157	1.0118	1.0352

Table 8-9 Six-step ahead forecast error comparison (YW and CLS) for INAR(1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$			
	$n=24$	$n=36$	$n=48$	$n=96$
$\alpha=0.1, \lambda=0.5$	0.9986	0.9995	0.9998	1.0000
$\alpha=0.5, \lambda=0.5$	0.9848	0.9938	0.9979	0.9993
$\alpha=0.9, \lambda=0.5$	0.9701	0.9786	0.9936	1.0139
$\alpha=0.1, \lambda=1$	0.9989	0.9997	0.9997	0.9999
$\alpha=0.5, \lambda=1$	0.9833	0.9936	0.9969	0.9994
$\alpha=0.9, \lambda=1$	0.9509	0.9723	0.9897	1.0138
$\alpha=0.1, \lambda=3$	0.9987	0.9995	0.9998	0.9999
$\alpha=0.5, \lambda=3$	0.9907	0.9918	0.9980	0.9994
$\alpha=0.9, \lambda=3$	0.9576	0.9679	0.9880	1.0175
$\alpha=0.1, \lambda=5$	0.9984	0.9996	0.9996	0.9999
$\alpha=0.5, \lambda=5$	0.9883	0.9952	0.9960	0.9992
$\alpha=0.9, \lambda=5$	0.9566	0.9738	0.9854	1.0101

Table 8-10 Three-step ahead forecast error comparison (YW and CLS) for INMA(1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$			
	$n=24$	$n=36$	$n=48$	$n=96$
$\beta=0.1, \lambda=0.5$	0.9940	0.9957	0.9980	0.9988
$\beta=0.5, \lambda=0.5$	0.9895	0.9909	0.9939	0.9967
$\beta=0.9, \lambda=0.5$	0.9811	0.9886	0.9918	0.9959
$\beta=0.1, \lambda=1$	0.9948	0.9964	0.9988	0.9984
$\beta=0.5, \lambda=1$	0.9934	0.9916	0.9929	0.9968
$\beta=0.9, \lambda=1$	0.9774	0.9851	0.9896	0.9954
$\beta=0.1, \lambda=3$	0.9947	0.9961	0.9968	0.9981
$\beta=0.5, \lambda=3$	0.9878	0.9940	0.9953	0.9963
$\beta=0.9, \lambda=3$	0.9951	0.9917	0.9932	0.9953
$\beta=0.1, \lambda=5$	0.9927	0.9957	0.9976	0.9985
$\beta=0.5, \lambda=5$	0.9940	0.9935	0.9948	0.9988
$\beta=0.9, \lambda=5$	0.9970	0.9998	0.9970	0.9978

Table 8-11 Six-step ahead forecast error comparison (YW and CLS) for INMA(1) series

Parameters	$\mathrm{MSE_{YW}}/\mathrm{MSE_{CLS}}$			
	$n=24$	$n=36$	$n=48$	$n=96$
$\beta=0.1, \lambda=0.5$	0.9927	0.9955	0.9970	0.9984
$\beta=0.5, \lambda=0.5$	0.9846	0.9946	0.9925	0.9973
$\beta=0.9, \lambda=0.5$	0.9798	0.9953	0.9942	0.9960
$\beta=0.1, \lambda=1$	0.9921	0.9970	0.9975	0.9989
$\beta=0.5, \lambda=1$	0.9907	0.9908	0.9939	0.9955
$\beta=0.9, \lambda=1$	0.9821	0.9844	0.9909	0.9945
$\beta=0.1, \lambda=3$	0.9929	0.9931	0.9968	0.9989
$\beta=0.5, \lambda=3$	0.9917	0.9911	0.9942	0.9963
$\beta=0.9, \lambda=3$	0.9919	0.9936	0.9953	0.9934
$\beta=0.1, \lambda=5$	0.9935	0.9954	0.9969	0.9980
$\beta=0.5, \lambda=5$	0.9919	0.9910	0.9967	0.9984
$\beta=0.9, \lambda=5$	1.0013	0.9958	0.9935	0.9980

For an INARMA(1,1) process, as can be seen from Table 8-12 and Table 8-13, CLS does not always produce better forecasts than YW. For high number of observations, the results are close.

Table 8-12 Three-step ahead forecast error comparison (YW and CLS) for INARMA(1,1) series

Parameters	$\mathrm{MSE}_{\mathrm{YW}}/\mathrm{MSE}_{\mathrm{CLS}}$			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	1.0301	1.0092	1.0051	1.0020
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.9986	0.9928	0.9977	0.9973
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9752	0.9976	0.9965	1.0009
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.8939	0.9184	0.9467	1.0027
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0217	1.0082	1.0100	1.0022
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9825	0.9949	0.9958	0.9968
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9531	1.0014	1.0013	1.0004
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.9139	1.0007	0.9974	1.0353
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.0147	1.0143	1.0072	1.0003
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9696	0.9956	0.9892	0.9921
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9635	0.9725	0.9923	0.9919
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.1665	1.0307	1.0383	1.0639

Table 8-13 Six-step ahead forecast error comparison (YW and CLS) for INARMA(1,1) series

Parameters	$\mathrm{MSE}_{\mathrm{YW}}/\mathrm{MSE}_{\mathrm{CLS}}$			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	1.0094	1.0088	1.0050	1.0006
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.9936	0.9933	0.9977	0.9975
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9720	0.9965	0.9975	0.9976
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.8577	0.8722	0.9262	0.9797
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0157	1.0127	1.0047	1.0010
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9735	0.9956	0.9964	0.9963
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9659	0.9947	0.9939	0.9959
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.9251	0.9431	0.9648	0.9977
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.0072	1.0193	1.0027	1.0017
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9688	0.9872	0.9930	0.9919
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9167	0.9718	0.9769	0.9907
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.1674	1.0402	1.0173	1.0190

For an INAR(1) process, the results of this section show that CLS produces better one-step ahead forecasts than YW for high autoregressive parameters. For lower autoregressive parameters, YW slightly outperform CLS. For an INMA(1) process, generally for high values of λ, YW is slightly better than CLS in terms of MSE of one-step ahead forecasts; but the opposite is true for lower values of λ. For an INARMA(1,1) process, CLS always produces better one-step ahead forecast than

YW using MSE.

However, the results of three-step and six-step ahead forecasts show that although the YW and CLS results are generally very close for all of the three INARMA processes, YW results are slightly better in many cases.

Therefore, based on the superior performance of CLS for one-step ahead forecasts, in the following sections where we compare the INARMA forecasts with those of benchmark methods, we use the CLS to estimate the parameters for one-step ahead INARMA forecasts. For three-step and six-step ahead forecasts, on the other hand, we use YW to estimate the parameters of the INARMA process.

8.5 Croston-SBA Categorization

Syntetos et al. (2005) compare Croston and SBA, based on MSE, to establish the areas that each method should be used over the other. The squared coefficient of variation (CV^2) of demand size and the average inter-demand interval (p) are used to identify the areas.

The coefficient of variation is defined by $\upsilon = \sigma/\mu_1^{'}$, where $\mu^{'}$ is the value of the mean measured from some arbitrary origin. It is estimated using the formula:

$$\hat{\upsilon} = \frac{\left(\frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_1^{'})^2\right)^{1/2}}{\frac{1}{n}\sum_{j=1}^{n} x_j}$$

Equation 8-1

where n is the sample size. The results show that for the smoothing parameter $\alpha = 0.2$, when $p > 1.31$, SBA is superior to Croston's method in terms of MSE. For $p \leq 1.31$, if $CV^2 > 0.47$ then $MSE_{Croston} > MSE_{SBA}$, but if $CV^2 \leq 0.47$ then Croston's method performs better in terms of MSE (Syntetos et al., 2005). This is shown in Figure 8-1. The cut-off values are slightly different for different smoothing parameters (Syntetos et al., 2005).

The Croston-SBA categorization is based on the assumption that demand occurs as an i.i.d. Bernoulli process. Therefore it is worth testing if it also holds for an i.i.d.

Poisson process. An INARMA(0,0) process produces such data series.

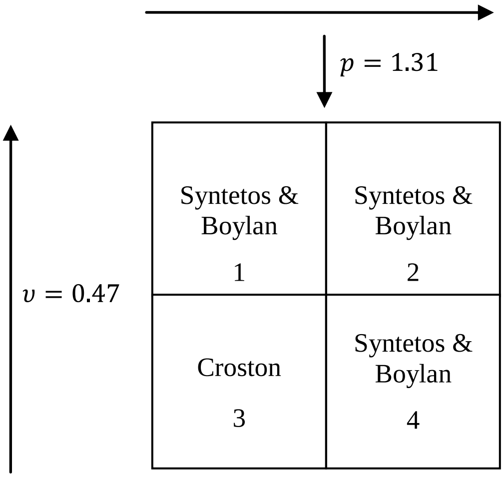

Figure 8-1 Cut-off values for Croston and SBA when $\alpha = 0.2$ (Syntetos et al., 2005)

The simulation results show that the squared coefficient of variation of demand size (CV^2) is always less than the cut-off value determined by Syntetos et al. (2005) (0.47). This is due to the Poisson assumption of demand. The p-values, however, vary below and beyond the cut-off value (1.31). Therefore, the demand series produced by an INARMA(0,0) process could belong to either region 3 or 4 in Figure 8-1.

The results of simulation confirm the Croston-SBA categorization for both cases of $\alpha = 0.2, 0.5$. Therefore, for $p \leq 1.31$ (or $\lambda \geq 3$), $\text{MSE}_{\text{Croston}} < \text{MSE}_{\text{SBA}}$ and for $p > 1.31$ (or $\lambda < 3$), $\text{MSE}_{\text{SBA}} < \text{MSE}_{\text{Croston}}$ when either all points in time or issue points are considered.

Although the Croston-SBA categorization is based on MSE, the results show that it generally holds for MASE as well. This was expected due to the similarities between the two error measures. However, there are some exceptions. For the case of $\lambda = 3$ where Croston's method should outperform SBA, $\text{MASE}_{\text{SBA}} < \text{MASE}_{\text{Croston}}$ for both cases of $\alpha = 0.2, 0.5$. Because the corresponding p-value and CV^2 of size for these cases are $p = 1.0538$ and $CV^2 = 0.2622$, these exceptions can be attributed to the nonlinear boundaries between region 3 and others in Figure 8-1 (Syntetos, 2001; Kostenko and Hyndman, 2006). This is an interesting finding because MASE is a relatively new measure and has been suggested for intermittent demand studies

(Hyndman, 2006).

The results show that when the number of observations increases, the advantage of SBA over Croston decreases. This is a new finding and this issue has not been discussed previously in the literature. It is shown is Table 8-14.

Table 8-14 The advantage of SBA over Croston for $\alpha = 0.2$ and all points in time

Parameters	$\mathrm{MSE_{SBA}} - \mathrm{MSE_{Croston}}$					$\mathrm{MASE_{SBA}} - \mathrm{MASE_{Croston}}$				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\lambda = 0.3$	-0.0186	-0.0085	-0.0051	-0.0025	-0.0025	-0.0507	-0.0431	-0.0392	-0.0355	-0.0342
$\lambda = 0.5$	-0.0168	-0.0084	-0.0061	-0.0051	-0.0052	-0.0245	-0.0197	-0.0176	-0.0175	-0.0174
$\lambda = 0.7$	-0.0173	-0.0100	-0.0082	-0.0079	-0.0077	-0.0116	-0.0076	-0.0067	-0.0069	-0.0063
$\lambda = 1$	-0.0173	-0.0118	-0.0111	-0.0107	-0.0102	-0.0152	-0.0131	-0.0134	-0.0130	-0.0131

Table 8-15 The advantage of SBA over Croston for $\alpha = 0.5$ and all points in time

Parameters	$\mathrm{MSE_{SBA}} - \mathrm{MSE_{Croston}}$					$\mathrm{MASE_{SBA}} - \mathrm{MASE_{Croston}}$				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\lambda = 0.3$	-0.0407	-0.0266	-0.0229	-0.0209	-0.0215	-0.1285	-0.1168	-0.1102	-0.1043	-0.1000
$\lambda = 0.5$	-0.0495	-0.0430	-0.0403	-0.0407	-0.0390	-0.0733	-0.0673	-0.0629	-0.0624	-0.0612
$\lambda = 0.7$	-0.0623	-0.0595	-0.0547	-0.0562	-0.0558	-0.0486	-0.0473	-0.0438	-0.0461	-0.0439
$\lambda = 1$	-0.0780	-0.0743	-0.0761	-0.0715	-0.0696	-0.0442	-0.0427	-0.0441	-0.0415	-0.0416

It can be seen from Table 8-14 that the advantage of SBA over Croston decreases when n increases.

The MSE of one step ahead forecast for a stationary mean model is:

$$\mathrm{MSE} = \mathrm{var(Estimates)} + \mathrm{Bias}^2 + \mathrm{var(Actual\ Demand)}$$

Equation 8-2

Syntetos (2001) assumes an infinite history for SES estimates:

$$\hat{Y}_t = \sum_{n=0}^{\infty} \alpha(1-\alpha)^n Y_{t-n}$$

Equation 8-3

and, therefore, Brown's expression for the variance of estimates is independent of the length of demand history, n:

$$\mathrm{var}(\hat{Y}_t) = \frac{\alpha}{2-\alpha}\mathrm{var}(Y_t)$$

Equation 8-4

However, when this assumption is relaxed, the finite representation of SES becomes (Graves, 1999):

$$\hat{Y}_t = \alpha Y_{t-1} + \alpha(1-\alpha)Y_{t-2} + \cdots + \alpha(1-\alpha)^{n-1}Y_{t-n+1} + (1-\alpha)^n Y_{t-n}$$

Equation 8-5

Therefore, the variance of the estimates produced by SES with a finite history is:

$$\mathrm{var}\left(\hat{Y}_t\right) = \frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\mathrm{var}(Y_t)$$

Equation 8-6

The variances of the exponentially smoothed size of demand and inter-demand interval with finite observations are then:

$$\mathrm{var}(z_t^{'}) = \frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\mathrm{var}(z_t)$$

Equation 8-7

$$\mathrm{var}(p_t^{'}) = \frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\mathrm{var}(p_t)$$

Equation 8-8

where $\mathrm{var}(z_t) = \sigma^2$ and $\mathrm{var}(p_t) = p(p-1)$, since the inter-demand interval follows the geometric distribution. The variance of the estimates produced by Croston's method is:

$$\mathrm{var}(Y_t^{'}) = \mathrm{var}\left(\frac{z_t^{'}}{p_t^{'}}\right)$$

Equation 8-9

The variance of the ratio of two independent variables is given by (Stuart and Ord, 1994):

$$\mathrm{var}\left(\frac{x}{y}\right) = \left(\frac{E(x)}{E(y)}\right)^2 \left[\frac{\mathrm{var}(x)}{[E(x)]^2} + \frac{\mathrm{var}(y)}{[E(y)]^2}\right]$$

Equation 8-10

The variance of the estimates produced by Croston's method with finite sample is therefore:

$$\mathrm{var}(Y_t^{'}) = \frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\left[\frac{(p-1)^2\mu^2}{p^4} + \frac{\sigma^2}{p^2}\right]$$

Equation 8-11

It can be seen that for $\alpha \in (0,1]$, as n increases, the above coefficient decreases until it reaches a limit of $[\alpha/(2-\alpha)]$. For high values of α, the limit is approached very quickly.

The difference between MSE of SBA and Croston's method is:

$$\mathrm{MSE_{SBA}} - \mathrm{MSE_{Croston}} = \left[\mathrm{var(Estimates_{SBA})} + \mathrm{Bias_{SBA}^2}\right] - \left[\mathrm{var(Estimates_{Croston})} + \mathrm{Bias_{Croston}^2}\right]$$

Equation 8-12

But when n increases, the bias of both Croston and SBA decreases and is close to zero (this has been confirmed by simulation results). So the difference is approximately:

$$\mathrm{MSE_{SBA}} - \mathrm{MSE_{Croston}} \approx \left[\left(1-\frac{\alpha}{2}\right)^2 - 1\right]\mathrm{var(Estimates_{Croston})}$$

Equation 8-13

From Equation 8-11 we have:

$$\mathrm{MSE_{SBA}} - \mathrm{MSE_{Croston}} \approx \left[\left(1-\frac{\alpha}{2}\right)^2 - 1\right]\left[\frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\right]\left[\frac{(p-1)^2\mu^2}{p^4} + \frac{\sigma^2}{p^2}\right]$$

Equation 8-14

As the results of Table 8-14 show, when $\alpha = 0.2$, the above coefficient decreases when n increases; therefore the difference between MSE of Croston and SBA also decreases. However, because the above coefficient reaches a limit of $\left(\frac{\alpha^2}{4} - \alpha\right)\left(\frac{\alpha}{2-\alpha}\right)$, it can be seen from Table 8-14 that the advantage of SBA over Croston does not change perceptibly when the number of observations is high.

For $\alpha = 0.5$, the results of Table 8-15 confirm that, as expected, the difference between MSE of Croston and SBA changes little with changes in n.

Although the Croston-SBA categorization is for i.i.d. demand, we have also tested it when demand is an INAR(1), INMA(1) or an INARMA(1,1) process. The results confirm that the Croston-SBA categorization generally holds for all of the above-mentioned processes. There is only one exception when there is an autoregressive component (either INAR(1) or INARMA(1,1)). For the INMA(1) case, there are two exceptions to the Croston-SBA categorization. The results are presented in Appendix 8.C.

Therefore, because the Croston-SBA categorization generally holds when the data is produced by any of the four INARMA processes, the best benchmark can be used to compete with INARMA forecasting methods.

8.6 INARMA vs Benchmark Methods

This research has suggested using INARMA models to forecast intermittent demand. In order to answer the last research question of "Do INARMA models provide more accurate forecasts for intermittent demand than non-optimal smoothing-based methods?", the performance of INARMA forecasts based on ME, MSE and MASE has been compared to that of benchmark methods. As previously mentioned, the benchmarks are Croston (Croston, 1972), SBA (Syntetos and Boylan, 2005) and SBJ (Shale et al., 2006) methods.

The first steps in the INARMA methodology are identification and estimation. These steps make INARMA more complicated than the benchmarks and result in two types of errors: error of identification and error of estimation. In order to investigate the effect of the identification error we first assume that the order of the model is known. The results are studied in section 8.6.1. Then we relax this assumption and examine the results for unknown model orders in section 8.6.2. Finally, the lead-time forecasts are compared in section 8.6.3.

8.6.1 INARMA with Known Order

In this section, we first compare the one-step ahead forecasts produced by each method (INARMA, Croston, SBA, and SBJ). The three-step and six-step ahead forecasts are then compared.

The results show that INARMA almost always produces the lowest MSE for all four processes (INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1)) when all points in time are considered. This is expected because the INARMA one-step ahead forecasts are MMSE forecasts and therefore when the demand follows an INARMA process, INARMA forecasts should outperform the benchmarks in terms of MSE.

The results also confirm that when only issue points are considered, the INARMA forecasts are biased (see Appendix 8.D). This is expected because the least squares criterion and therefore the CLS estimates are developed for the case where parameters are updated each period regardless of the demand being positive or zero.

As parameters need to be estimated for INARMA models, with an increase in the number of observations, the forecasts' accuracy increases.

The degree of improvement over benchmarks using the MSE measure, when all points in time are considered, is shown in Table 8-16 to Table 8-21. It should be noted that in some tables α is used in two different ways: when it is below the benchmark methods, it is the smoothing parameter for that specific method and when it is in the first column of the table, it is the autoregressive parameter of the INARMA process.

Table 8-16 One-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ for INARMA(0,0) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\lambda=0.3$	0.8691	0.9114	0.9150	0.9473	0.9711	0.9731	0.9740	0.9894	0.9904	0.9794	0.9872	0.9875
$\lambda=0.5$	0.9502	0.9773	0.9789	0.9747	0.9895	0.9904	0.9762	0.9870	0.9876	0.9675	0.9769	0.9772
$\lambda=0.7$	0.9694	0.9900	0.9917	0.9732	0.9857	0.9862	0.9706	0.9812	0.9814	0.9587	0.9688	0.9690
$\lambda=1$	0.9749	0.9892	0.9894	0.9688	0.9789	0.9788	0.9575	0.9670	0.9668	0.9478	0.9572	0.9569
$\lambda=3$	0.9596	0.9525	0.9484	0.9424	0.9369	0.9331	0.9323	0.9257	0.9218	0.9208	0.9136	0.9096
$\lambda=5$	0.9498	0.9255	0.9178	0.9367	0.9112	0.9034	0.9269	0.9048	0.8974	0.9126	0.8889	0.8814
$\lambda=20$	0.9529	0.8212	0.7940	0.9315	0.8023	0.7756	0.9281	0.7990	0.7721	0.9124	0.7851	0.7586

The results show that the improvement increases when more observations are available for higher values of λ ($\lambda \geq 0.7$). With more observations, the accuracy of parameters' estimates and therefore forecasts of INARMA and benchmark methods become more accurate. However, the results suggest that the accuracy of INARMA forecasts improves at a faster rate than the benchmarks.

The simulation results also show that for INARMA(0,0) and INMA(1) processes, the improvement over benchmarks is not considerable. However, with the presence of an autoregressive component, as in the INAR(1) and INARMA(1,1) cases, the improvement is considerable for the cases in which the autoregressive parameter is high ($\alpha = 0.9$).

As can be seen from Table 8-16, when $\lambda = 20$, the improvement of MSE of INARMA over SBA and SBJ is very high. This is because these methods are designed for highly intermittent demand, but when $\lambda = 20$, the demand is barely intermittent and the methods do not reduce to SES. In this case, Croston's method is equivalent to SES.

Table 8-17 One-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ for INMA(1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9421	0.9733	0.9755	0.9634	0.9816	0.9829	0.9715	0.9833	0.9837	0.9583	0.9704	0.9709
$\beta = 0.5, \lambda = 0.5$	0.9124	0.9470	0.9490	0.9395	0.9645	0.9660	0.9400	0.9607	0.9620	0.9282	0.9483	0.9495
$\beta = 0.9, \lambda = 0.5$	0.9124	0.9543	0.9569	0.9227	0.9503	0.9521	0.9273	0.9503	0.9515	0.9118	0.9352	0.9367
$\beta = 0.1, \lambda = 1$	0.9739	0.9871	0.9871	0.9629	0.9752	0.9751	0.9566	0.9673	0.9671	0.9426	0.9532	0.9530
$\beta = 0.5, \lambda = 1$	0.9872	1.0027	1.0024	0.9685	0.9793	0.9785	0.9550	0.9674	0.9669	0.9425	0.9545	0.9539
$\beta = 0.9, \lambda = 1$	1.0070	1.0146	1.0129	0.9849	0.9945	0.9931	0.9750	0.9839	0.9825	0.9628	0.9706	0.9690
$\beta = 0.1, \lambda = 3$	0.9719	0.9611	0.9563	0.9587	0.9490	0.9442	0.9500	0.9390	0.9341	0.9334	0.9249	0.9204
$\beta = 0.5, \lambda = 3$	1.0174	0.9948	0.9870	1.0035	0.9805	0.9728	0.9954	0.9730	0.9655	0.9763	0.9536	0.9461
$\beta = 0.9, \lambda = 3$	1.0720	1.0282	1.0166	1.0657	1.0262	1.0150	1.0405	1.0045	0.9937	1.0300	0.9917	0.9809
$\beta = 0.1, \lambda = 5$	0.9788	0.9474	0.9383	0.9499	0.9252	0.9169	0.9445	0.9168	0.9082	0.9308	0.9028	0.8944
$\beta = 0.5, \lambda = 5$	1.0318	0.9787	0.9648	1.0241	0.9679	0.9537	1.0019	0.9461	0.9322	0.9866	0.9376	0.9244
$\beta = 0.9, \lambda = 5$	1.0969	1.0211	1.0025	1.0777	1.0051	0.9867	1.0635	0.9917	0.9736	1.0345	0.9622	0.9445

When data is produced by an INARMA(0,0) or an INMA(1) process, the results of INARMA forecasts are only compared to those of Croston, SBA and SBJ with smoothing parameter 0.2. But for INAR(1) and INARMA(1,1) processes where an autoregressive component is present, the benchmark methods with smoothing parameter 0.5 are also included in comparisons (Table 8-19 and Table 8-21). This is because when the autoregressive parameter is high, the benchmark methods with higher smoothing parameter produce better forecasts than those with smoothing parameter 0.2.

Table 8-18 One-step ahead $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ with smoothing parameter 0.2 for INAR(1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \lambda = 0.5$	0.9752	1.0033	1.0026	0.9868	1.0061	1.0074	0.9747	0.9877	0.9884	0.9609	0.9726	0.9733
$\alpha = 0.5, \lambda = 0.5$	0.7962	0.8190	0.8212	0.7715	0.7929	0.7941	0.7688	0.7852	0.7858	0.7408	0.7582	0.7590
$\alpha = 0.9, \lambda = 0.5$	0.5848	0.5134	0.4988	0.5333	0.4666	0.4536	0.5251	0.4684	0.4559	0.4917	0.4396	0.4280
$\alpha = 0.1, \lambda = 1$	0.9971	1.0153	1.0158	0.9746	0.9860	0.9858	0.9636	0.9725	0.9721	0.9433	0.9532	0.9529
$\alpha = 0.5, \lambda = 1$	0.8845	0.8890	0.8870	0.8280	0.8308	0.8288	0.8122	0.8145	0.8124	0.7924	0.7962	0.7943
$\alpha = 0.9, \lambda = 1$	0.5825	0.4598	0.4379	0.5478	0.4285	0.4078	0.5275	0.4196	0.3997	0.5068	0.4054	0.3865
$\alpha = 0.1, \lambda = 3$	0.9922	0.9794	0.9742	0.9621	0.9539	0.9494	0.9538	0.9433	0.9385	0.9317	0.9222	0.9176
$\alpha = 0.5, \lambda = 3$	0.9342	0.8928	0.8815	0.8807	0.8408	0.8301	0.8586	0.8180	0.8073	0.8356	0.7960	0.7857
$\alpha = 0.9, \lambda = 3$	0.5985	0.3502	0.3175	0.5379	0.3104	0.2816	0.5215	0.3001	0.2723	0.4984	0.2848	0.2584
$\alpha = 0.1, \lambda = 5$	0.9982	0.9644	0.9550	0.9675	0.9353	0.9261	0.9515	0.9224	0.9135	0.9325	0.9031	0.8943
$\alpha = 0.5, \lambda = 5$	0.9170	0.8396	0.8217	0.8934	0.8194	0.8019	0.8586	0.7846	0.7677	0.8383	0.7676	0.7511
$\alpha = 0.9, \lambda = 5$	0.5881	0.2457	0.2165	0.5374	0.2341	0.2067	0.5208	0.2322	0.2052	0.5022	0.2190	0.1934

Table 8-19 One-step ahead $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ with smoothing parameter 0.5 for INAR(1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \lambda = 0.5$	0.8996	0.9844	1.0063	0.8776	0.9584	0.9960	0.8577	0.9340	0.9680	0.8445	0.9192	0.9505
$\alpha = 0.5, \lambda = 0.5$	0.7901	0.8582	0.7930	0.7459	0.8201	0.7663	0.7356	0.8014	0.7470	0.7034	0.7688	0.7252
$\alpha = 0.9, \lambda = 0.5$	0.9191	0.3860	0.2257	0.8746	0.3646	0.2094	0.8416	0.3604	0.2104	0.8066	0.3418	0.1975
$\alpha = 0.1, \lambda = 1$	0.8912	0.9544	0.9629	0.8619	0.9153	0.9243	0.8548	0.9036	0.9074	0.8318	0.8836	0.8915
$\alpha = 0.5, \lambda = 1$	0.9069	0.8851	0.7594	0.8500	0.8284	0.7064	0.8316	0.8128	0.6949	0.8058	0.7894	0.6785
$\alpha = 0.9, \lambda = 1$	0.9332	0.2532	0.1451	0.8807	0.2362	0.1349	0.8583	0.2296	0.1316	0.8248	0.2256	0.1294
$\alpha = 0.1, \lambda = 3$	0.8630	0.8141	0.7700	0.8310	0.7906	0.7558	0.8293	0.7850	0.7434	0.8105	0.7676	0.7264
$\alpha = 0.5, \lambda = 3$	0.9728	0.7309	0.5562	0.9190	0.6883	0.5230	0.8957	0.6672	0.5067	0.8703	0.6505	0.4946
$\alpha = 0.9, \lambda = 3$	0.9337	0.1106	0.0630	0.8746	0.1003	0.0569	0.8514	0.0972	0.0552	0.8186	0.0922	0.0523
$\alpha = 0.1, \lambda = 5$	0.8603	0.7393	0.6599	0.8312	0.7100	0.6363	0.8165	0.6998	0.6270	0.8010	0.6854	0.6142
$\alpha = 0.5, \lambda = 5$	0.9588	0.5936	0.4247	0.9311	0.5748	0.4116	0.8952	0.5533	0.3949	0.8720	0.5369	0.3847
$\alpha = 0.9, \lambda = 5$	0.9364	0.0642	0.0362	0.8728	0.0620	0.0350	0.8586	0.0612	0.0346	0.8171	0.0578	0.0326

The degree of improvement over benchmarks, using the MASE measure, is shown in Appendix 8.E. The results show that for INARMA(0,0) and INMA(1) processes, the improvement over benchmarks, in terms of MASE, is not considerable. But for INAR(1) and INARMA(1,1) processes, the improvement is high for the cases in which the autoregressive parameter is high. This confirms the results using MSE.

Table 8-20 One-step ahead $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ with smoothing parameter 0.2 for INARMA(1,1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.9881	1.0188	1.0202	0.9798	1.0007	1.0021	0.9716	0.9882	0.9892	0.9516	0.9667	0.9674
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.8263	0.8614	0.8633	0.8117	0.8382	0.8400	0.8272	0.8483	0.8495	0.8113	0.8311	0.8322
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.7641	0.7846	0.7846	0.7271	0.7439	0.7443	0.7000	0.7128	0.7130	0.6491	0.6638	0.6641
$\alpha=0.9,\beta=0.1,\lambda=0.5$	0.5621	0.4725	0.4577	0.5314	0.4674	0.4542	0.4982	0.4397	0.4276	0.4740	0.4208	0.4091
$\alpha=0.1,\beta=0.1,\lambda=1$	1.0022	1.0176	1.0176	0.9872	1.0013	1.0012	0.9657	0.9782	0.9780	0.9410	0.9530	0.9528
$\alpha=0.1,\beta=0.9,\lambda=1$	0.9076	0.9143	0.9127	0.8657	0.8765	0.8754	0.8613	0.8670	0.8655	0.8523	0.8595	0.8581
$\alpha=0.5,\beta=0.5,\lambda=1$	0.8216	0.8156	0.8119	0.7701	0.7676	0.7645	0.7533	0.7488	0.7456	0.7082	0.7073	0.7045
$\alpha=0.9,\beta=0.1,\lambda=1$	0.5736	0.4402	0.4180	0.5206	0.4057	0.3860	0.4975	0.3911	0.3723	0.4610	0.3665	0.3492
$\alpha=0.1,\beta=0.1,\lambda=5$	1.0117	0.9720	0.9613	0.9802	0.9453	0.9352	0.9653	0.9306	0.9206	0.9442	0.9096	0.8998
$\alpha=0.1,\beta=0.9,\lambda=5$	0.9149	0.8425	0.8257	0.8857	0.8149	0.7987	0.8560	0.7915	0.7762	0.8482	0.7853	0.7701
$\alpha=0.5,\beta=0.5,\lambda=5$	0.8271	0.7310	0.7106	0.7831	0.6908	0.6713	0.7541	0.6713	0.6531	0.7249	0.6439	0.6262
$\alpha=0.9,\beta=0.1,\lambda=5$	0.5341	0.2424	0.2140	0.5230	0.2214	0.1951	0.4948	0.2153	0.1901	0.8477	0.8073	0.0154

Table 8-21 One-step ahead $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ with smoothing parameter 0.5 for INARMA(1,1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.9038	1.0019	1.0255	0.8647	0.9572	0.9917	0.8475	0.9364	0.9709	0.8258	0.9106	0.9474
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.7641	0.8583	0.8595	0.7157	0.8114	0.8250	0.7357	0.8203	0.8208	0.7211	0.8043	0.8037
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.7960	0.8529	0.7388	0.7567	0.7983	0.6949	0.7268	0.7625	0.6618	0.6713	0.7108	0.6191
$\alpha=0.9,\beta=0.1,\lambda=0.5$	0.9343	0.3702	0.2060	0.8703	0.3592	0.2061	0.8319	0.3426	0.1959	0.7895	0.3213	0.1843
$\alpha=0.1,\beta=0.1,\lambda=1$	0.8982	0.9543	0.9513	0.8780	0.9369	0.9357	0.8621	0.9150	0.9112	0.8355	0.8885	0.8884
$\alpha=0.1,\beta=0.9,\lambda=1$	0.8859	0.8900	0.7964	0.8257	0.8452	0.7746	0.8359	0.8396	0.7565	0.8209	0.8294	0.7517
$\alpha=0.5,\beta=0.5,\lambda=1$	0.9256	0.8391	0.6537	0.8635	0.7895	0.6188	0.8433	0.7673	0.6014	0.7877	0.7255	0.5714
$\alpha=0.9,\beta=0.1,\lambda=1$	0.9236	0.2359	0.1346	0.8665	0.2240	0.1270	0.8322	0.2161	0.1228	0.7864	0.2054	0.1164
$\alpha=0.1,\beta=0.1,\lambda=5$	0.8940	0.7391	0.6392	0.8696	0.7208	0.6244	0.8625	0.7127	0.6135	0.8371	0.6939	0.6003
$\alpha=0.1,\beta=0.9,\lambda=5$	0.9156	0.5979	0.4408	0.8931	0.5789	0.4277	0.8616	0.5651	0.4186	0.8536	0.5581	0.4132
$\alpha=0.5,\beta=0.5,\lambda=5$	0.9496	0.4851	0.3221	0.8921	0.4558	0.3031	0.8605	0.4491	0.2998	0.8230	0.4266	0.2851
$\alpha=0.9,\beta=0.1,\lambda=5$	0.8881	0.0630	0.0357	0.8565	0.0576	0.0325	0.8252	0.0568	0.0320	0.9035	0.6340	0.5079

The results of comparing INARMA with benchmark methods for three-step ahead forecasts are given in Table 8-22 to Table 8-27 (See Appendix 8.F for the six-step ahead results). The results of comparing the h-step ahead forecasts ($h = 3, 6$) of INARMA with benchmarks for INARMA(0,0) are very close to the results of one-step ahead forecasts.

Table 8-22 Three-step ahead $MSE_{INARMA}/MSE_{Benchmark}$ for INARMA(0,0) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\lambda=0.3$	0.8459	0.8930	0.8980	0.9447	0.9692	0.9714	0.9751	0.9886	0.9898	0.9826	0.9901	0.9904
$\lambda=0.5$	0.9458	0.9761	0.9785	0.9756	0.9910	0.9919	0.9735	0.9846	0.9851	0.9665	0.9757	0.9759
$\lambda=0.7$	0.9693	0.9909	0.9924	0.9751	0.9868	0.9871	0.9720	0.9824	0.9827	0.9586	0.9686	0.9687
$\lambda=1$	0.9758	0.9901	0.9903	0.9686	0.9791	0.9789	0.9644	0.9724	0.9720	0.9475	0.9564	0.9561
$\lambda=3$	0.9578	0.9475	0.9432	0.9411	0.9346	0.9307	0.9305	0.9222	0.9182	0.9196	0.9137	0.9100
$\lambda=5$	0.9522	0.9320	0.9249	0.9389	0.9149	0.9073	0.9263	0.9029	0.8954	0.9118	0.8908	0.8836
$\lambda=20$	0.9549	0.8234	0.7960	0.9331	0.8011	0.7740	0.9265	0.7968	0.7699	0.9152	0.7911	0.7649

Table 8-23 Three-step ahead $MSE_{INARMA}/MSE_{Benchmark}$ for INMA(1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\beta=0.1,\lambda=0.5$	0.9359	0.9747	0.9782	0.9738	0.9916	0.9928	0.9660	0.9812	0.9820	0.9593	0.9711	0.9716
$\beta=0.5,\lambda=0.5$	0.9514	0.9890	0.9922	0.9541	0.9782	0.9798	0.9503	0.9713	0.9726	0.9280	0.9474	0.9486
$\beta=0.9,\lambda=0.5$	0.9481	0.9968	1.0009	0.9423	0.9752	0.9776	0.9297	0.9595	0.9615	0.9142	0.9367	0.9380
$\beta=0.1,\lambda=1$	0.9727	0.9880	0.9882	0.9664	0.9784	0.9783	0.9530	0.9666	0.9667	0.9357	0.9470	0.9468
$\beta=0.5,\lambda=1$	0.9614	0.9816	0.9820	0.9430	0.9580	0.9579	0.9274	0.9431	0.9431	0.9066	0.9214	0.9213
$\beta=0.9,\lambda=1$	0.9597	0.9746	0.9741	0.9269	0.9385	0.9377	0.9113	0.9312	0.9312	0.8884	0.9023	0.9017
$\beta=0.1,\lambda=3$	0.9542	0.9490	0.9449	0.9369	0.9290	0.9246	0.9246	0.9208	0.9169	0.9097	0.9042	0.9001
$\beta=0.5,\lambda=3$	0.9426	0.9304	0.9246	0.9146	0.9060	0.9007	0.8970	0.8871	0.8818	0.8797	0.8728	0.8677
$\beta=0.9,\lambda=3$	0.9444	0.9307	0.9238	0.9052	0.8900	0.8830	0.8915	0.8789	0.8723	0.8626	0.8496	0.8431
$\beta=0.1,\lambda=5$	0.9535	0.9341	0.9265	0.9339	0.9077	0.8996	0.9143	0.8929	0.8853	0.9010	0.8786	0.8710
$\beta=0.5,\lambda=5$	0.9401	0.9164	0.9067	0.9199	0.8872	0.8769	0.8997	0.8650	0.8546	0.8741	0.8418	0.8319
$\beta=0.9,\lambda=5$	0.9493	0.9035	0.8903	0.9075	0.8631	0.8503	0.8929	0.8528	0.8406	0.8618	0.8220	0.8101

Table 8-24 Three-step ahead $MSE_{INARMA}/MSE_{Benchmark}$ with smoothing parameter 0.2 for INAR(1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\alpha=0.1,\lambda=0.5$	0.9779	1.0054	1.0076	0.9766	0.9924	0.9934	0.9688	0.9841	0.9850	0.9538	0.9672	0.9678
$\alpha=0.5,\lambda=0.5$	1.0023	1.0387	1.0413	0.9757	1.0063	1.0083	0.9403	0.9708	0.9728	0.8924	0.9209	0.9228
$\alpha=0.9,\lambda=0.5$	0.9722	0.9494	0.9365	0.9039	0.8583	0.8446	0.8577	0.8218	0.8098	0.7963	0.7682	0.7574
$\alpha=0.1,\lambda=1$	0.9789	0.9993	1.0002	0.9623	0.9752	0.9752	0.9497	0.9607	0.9605	0.9347	0.9469	0.9469
$\alpha=0.5,\lambda=1$	1.0005	1.0231	1.0231	0.9571	0.9759	0.9757	0.9256	0.9445	0.9444	0.8840	0.9012	0.9009
$\alpha=0.9,\lambda=1$	0.9861	0.8725	0.8463	0.9031	0.8023	0.7786	0.8624	0.7758	0.7537	0.8057	0.7239	0.7029
$\alpha=0.1,\lambda=3$	0.9586	0.9553	0.9512	0.9445	0.9363	0.9318	0.9253	0.9201	0.9161	0.9050	0.8995	0.8955
$\alpha=0.5,\lambda=3$	0.9964	0.9755	0.9670	0.9330	0.9170	0.9095	0.9118	0.8958	0.8884	0.8730	0.8639	0.8573
$\alpha=0.9,\lambda=3$	0.9680	0.6638	0.6176	0.8894	0.6050	0.5624	0.8873	0.6099	0.5665	0.8018	0.5559	0.5167
$\alpha=0.1,\lambda=5$	0.9576	0.9332	0.9251	0.9296	0.9014	0.8931	0.9172	0.8933	0.8853	0.9035	0.8787	0.8707
$\alpha=0.5,\lambda=5$	1.0170	0.9605	0.9454	0.9463	0.8926	0.8779	0.9114	0.8690	0.8558	0.8659	0.8225	0.8096
$\alpha=0.9,\lambda=5$	0.9575	0.5535	0.5009	0.8854	0.8709	0.8602	0.8585	0.4990	0.4524	0.7998	0.4644	0.4210

For an INMA(1) process, the performance of INARMA compared to benchmark methods is improved for *h*-step ahead forecasts compared to one-step ahead forecasts. This could be because the benchmark methods use the same forecast as one-step ahead forecast, but the INMA(1) method updates the forecasts.

Table 8-25 Three-step ahead $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.5 for INAR(1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \lambda = 0.5$	0.8885	0.9715	0.9774	0.8680	0.9432	0.9476	0.8537	0.9305	0.9356	0.8285	0.9070	0.9135
$\alpha = 0.5, \lambda = 0.5$	0.9040	1.0122	1.0132	0.8470	0.9572	0.9595	0.8166	0.9260	0.9295	0.7684	0.8704	0.8741
$\alpha = 0.9, \lambda = 0.5$	1.0808	0.7588	0.5810	1.0428	0.6862	0.5234	0.9866	0.6757	0.5192	0.9210	0.6331	0.4861
$\alpha = 0.1, \lambda = 1$	0.8477	0.9174	0.9078	0.8337	0.8935	0.8821	0.8181	0.8761	0.8642	0.8076	0.8658	0.8543
$\alpha = 0.5, \lambda = 1$	0.8654	0.9204	0.8876	0.8261	0.8750	0.8432	0.8023	0.8504	0.8206	0.7673	0.8105	0.7808
$\alpha = 0.9, \lambda = 1$	1.1085	0.5490	0.3835	1.0376	0.5089	0.3552	1.0002	0.4991	0.3480	0.9346	0.4595	0.3190
$\alpha = 0.1, \lambda = 3$	0.8006	0.7790	0.7181	0.7972	0.7646	0.7029	0.7724	0.7476	0.6896	0.7575	0.7331	0.6762
$\alpha = 0.5, \lambda = 3$	0.8384	0.7401	0.6411	0.7946	0.7033	0.6100	0.7727	0.6852	0.5947	0.7383	0.6629	0.5761
$\alpha = 0.9, \lambda = 3$	1.0951	0.2602	0.1617	1.0240	0.2353	0.1460	1.0154	0.2317	0.1432	0.9261	0.2125	0.1313
$\alpha = 0.1, \lambda = 5$	0.7956	0.7032	0.6145	0.7698	0.6751	0.5888	0.7597	0.6681	0.5823	0.7511	0.6579	0.5729
$\alpha = 0.5, \lambda = 5$	0.8714	0.6602	0.5333	0.8028	0.6011	0.4821	0.7715	0.5927	0.4782	0.7308	0.5574	0.4492
$\alpha = 0.9, \lambda = 5$	1.1015	0.1672	0.0996	1.0322	0.7275	0.5586	1.0000	0.1529	0.0914	0.9226	0.1424	0.0852

Table 8-26 Three-step ahead $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.2 for INARMA(1,1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9621	0.9993	0.9992	0.9806	0.9994	1.0006	0.9730	0.9897	0.9907	0.9526	0.9673	0.9680
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.9515	0.9996	1.0031	0.9404	0.9728	0.9753	0.9278	0.9577	0.9597	0.8986	0.9229	0.9244
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9945	1.0227	1.0239	0.9531	0.9855	0.9875	0.9317	0.9631	0.9651	0.8928	0.9228	0.9247
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.3727	1.2345	1.2092	1.5485	1.4763	1.4539	1.5901	1.5111	1.4876	1.7403	1.6794	1.6553
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9927	1.0168	1.0178	0.9670	0.9810	0.9810	0.9561	0.9699	0.9698	0.9220	0.9359	0.9360
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9532	0.9718	0.9715	0.9177	0.9347	0.9344	0.9064	0.9227	0.9224	0.8833	0.8989	0.8985
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9933	1.0106	1.0097	0.9427	0.9644	0.9641	0.9201	0.9347	0.9336	0.8862	0.9034	0.9027
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.3867	1.2511	1.2147	1.5431	1.3756	1.3331	1.5887	1.4125	1.3700	1.7368	1.5550	1.5095
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9832	0.9584	0.9495	0.9375	0.9126	0.9041	0.9181	0.8978	0.8898	0.8928	0.8704	0.8625
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9710	0.9245	0.9105	0.9130	0.8713	0.8584	0.8894	0.8487	0.8365	0.8556	0.8185	0.8068
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	1.0173	0.9483	0.9303	0.9598	0.9053	0.8889	0.9245	0.8736	0.8579	0.8803	0.8285	0.8133
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.7356	0.9574	0.8627	1.6699	0.9654	0.8737	1.7683	1.0071	0.9128	1.7151	0.9770	0.8836

For INAR(1) and INARMA(1,1) processes, the performance of INARMA over the benchmark methods is improved compared to the one-step ahead case when the autoregressive parameter is low. But, as discussed in chapter 6, when the autoregressive parameter is high, the fact that the forecasts converge to the mean of the process results in poor forecasts compared to the one-step ahead case. As explained in chapter 6, some authors suggest using different models for different horizons in order to improve forecast accuracy (Cox, 1961; Tiao and Xu, 1993; Kang, 2003).

Table 8-27 Three-step ahead $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.5 for INARMA(1,1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.8645	0.9711	1.0200	0.8670	0.9500	0.9861	0.8402	0.9282	0.9735	0.8223	0.9071	0.9479
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8371	0.9597	1.0175	0.7945	0.9047	0.9622	0.7757	0.8869	0.9431	0.7477	0.8483	0.9002
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.8855	0.9781	0.9853	0.8377	0.9372	0.9549	0.8185	0.9158	0.9341	0.7771	0.8723	0.8951
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.5506	0.9825	0.6679	1.8084	1.2198	0.8406	1.8603	1.2357	0.8578	2.0440	1.3820	0.9572
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.8588	0.9342	0.9719	0.8309	0.8943	0.9200	0.8203	0.8794	0.9068	0.7889	0.8504	0.8796
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.7944	0.8515	0.8782	0.7608	0.8154	0.8461	0.7513	0.8033	0.8338	0.7334	0.7820	0.8109
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8670	0.9014	0.8896	0.8186	0.8594	0.8564	0.7895	0.8212	0.8123	0.7642	0.7988	0.7923
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.5546	0.7857	0.5145	1.7618	0.8418	0.5498	1.8280	0.8853	0.5789	2.0404	0.9911	0.6449
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.8070	0.7083	0.6504	0.7714	0.6736	0.6199	0.7510	0.6620	0.6146	0.7323	0.6432	0.5955
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.7803	0.6129	0.5297	0.7334	0.5807	0.5030	0.7137	0.5707	0.4969	0.6856	0.5474	0.4774
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8725	0.6197	0.4987	0.8124	0.5889	0.4761	0.7889	0.5699	0.4608	0.7503	0.5401	0.4354
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.9454	0.2820	0.1644	1.9418	0.2908	0.1712	2.0688	0.3114	0.1832	1.9978	0.2951	0.1731

8.6.2 INARMA with Unknown Order

It was assumed in section 8.6.1 that the order of the INARMA process is known. However, in reality this is not the case and the autoregressive and moving average orders of the model need to be identified. As discussed in chapter 4, when simulating a high number of replications, automated approaches such as AIC and BIC should be used for identification rather than sample autocorrelation (SACF) and sample partial autocorrelation functions (SPACF).

In chapter 4, two identification procedures were suggested. A two-stage identification procedure is based on using the Ljung-Box test to distinguish between INARMA(0,0) and other processes and then using the AIC (or AIC_C for small sample sizes) to select among the other possible INARMA processes. On the other hand, the one-stage method only uses the AIC to select among all possible INARMA models including INARMA(0,0).

For simplicity, we first assume that data can be produced by either an INARMA(0,0) or an INAR(1) process. The results of identification among these two processes are summarized in section 8.6.2.1. As an alternative approach to identification, we suggest that the model with the highest order in the set of models (in this case INAR(1)) can be used for forecasting. The results are presented in section 8.6.2.2.

Then we assume that data can be produced by any of the four processes. The results of identification based using two-stage and one-stage identification procedures are presented in section 8.6.2.3. The most-general-model approach is also tested and the results are analyzed in section 8.6.2.4. The results of treating all models as INAR(1) are compared to the benchmark methods in section 8.6.2.5.

8.6.2.1 Identification among Two Processes

In this section, it is assumed that data is either INARMA(0,0) or INAR(1). As suggested by Jung and Tremayne (2003), in order to distinguish between the INARMA(0,0) (or an i.i.d. Poisson process) and INAR(1), we test if the data show a significant serial dependence or not.

This is done using a portmanteau test called the Ljung-Box test explained in section 4.2.3. The test is based on a Q^*-statistic given by:

$$Q^* = n(n+2)\sum_{j=1}^{k}(n-j)^{-1}\hat{\rho}_j^2$$

Equation 8-15

where n is the sample size, k is the number of autocorrelation lags included in the statistic (we assume $k = 10$), $\hat{\rho}_j$ is the sample autocorrelation at lag j. The Q^* statistic can be used when a univariate model is fitted to a time series. It can be used as a lack-of-fit test for a departure from randomness. Under the null hypothesis that the model fit is adequate, the test statistic is asymptotically chi-square distributed. Results are presented for a significance level of 0.05.

The results in terms of percentage of series for which the model is correctly identified are summarized in Table 8-28 and Table 8-29 for both Ljung-Box and AIC identification procedures. The results of Table 8-28 show identification with the

Ljung-Box test provides better results than with the AIC for INARMA(0,0) series.

Comparing the results of Table 8-28 to the results by Jung and Tremayne (2003) shows that the Q^* statistic produces similar results to those suggested in their study.

The simulation results show that for those cases where an INARMA(0,0) is misidentified as an INAR(1) process, the estimated autoregressive parameter is close to zero.

Table 8-28 The percentage of correct identification for INARMA(0,0) series

Parameters	Ljung-Box				AIC			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\lambda = 0.3$	94.42	95.53	97.10	95.00	91.00	89.60	87.70	81.50
$\lambda = 0.5$	93.35	94.50	94.60	95.60	92.20	88.30	83.70	78.70
$\lambda = 0.7$	93.16	94.11	96.00	95.60	91.80	87.80	83.10	74.80
$\lambda = 1$	93.64	94.25	94.60	95.30	91.10	83.20	80.40	71.40
$\lambda = 3$	95.00	92.50	94.20	95.20	83.10	79.60	74.70	69.70
$\lambda = 5$	92.80	93.70	93.80	94.60	84.20	77.20	74.40	68.80
$\lambda = 20$	93.20	94.20	93.60	94.80	83.70	75.90	72.40	68.50

It can be seen from Table 8-29 that, for an INAR(1) process, when the autoregressive parameter is small ($\alpha = 0.1$), the model is often misidentified as INARMA(0,0). The AIC is always better than the Ljung-Box method. For high values of α, the correct model in identified by the AIC in most cases. Both identification methods perform better when more observations are available. For high values of α ($\alpha = 0.9$) and n ($n = 96$) the two identification methods are close.

Table 8-29 The percentage of correct identification for INAR(1) series

Parameters	Ljung-Box				AIC			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \lambda = 0.5$	5.90	6.00	5.10	5.40	13.70	19.90	28.40	47.20
$\alpha = 0.5, \lambda = 0.5$	15.40	23.90	34.30	71.60	49.00	73.60	86.00	98.60
$\alpha = 0.9, \lambda = 0.5$	34.20	65.50	84.80	100.00	90.00	98.50	99.70	100.00
$\alpha = 0.1, \lambda = 1$	6.50	7.20	6.90	6.30	14.90	26.10	35.40	52.20
$\alpha = 0.5, \lambda = 1$	16.30	22.60	35.40	73.90	57.70	80.70	90.10	99.70
$\alpha = 0.9, \lambda = 1$	35.10	63.30	86.50	99.80	89.80	98.30	100.00	100.00
$\alpha = 0.1, \lambda = 3$	9.00	5.90	7.20	6.30	22.30	33.90	39.10	57.00
$\alpha = 0.5, \lambda = 3$	14.50	23.90	34.40	71.30	62.70	83.40	92.50	99.60
$\alpha = 0.9, \lambda = 3$	34.50	66.00	83.50	99.90	91.10	98.70	99.80	100.00
$\alpha = 0.1, \lambda = 5$	7.40	6.50	4.80	7.70	22.80	37.10	40.10	59.40
$\alpha = 0.5, \lambda = 5$	16.80	24.40	35.30	72.50	67.00	86.60	93.20	99.70
$\alpha = 0.9, \lambda = 5$	33.90	64.40	83.60	99.70	91.10	98.50	99.70	100.00

Because the INARMA(0,0) process is correctly identified in most of the cases, the forecast accuracy of the INARMA(0,0) with identification is close to that of the case when the order is known. The same is true for an INAR(1) process with high values of α and n. As a result, the performance of these two processes compared to benchmarks is similar to what was discussed in section 8.6.1.

The accuracy results of INAR(1) forecasts using ME, MSE and MASE for the two identification methods are shown in Table 8-30. Similar results for the case that the order of the model is known are provided in Table 8-31 for comparison.

The results of Table 8-30 show that the AIC produces better forecasts than the Ljung-Box method in most of the cases, except for the case where $\alpha = 0.1$ and λ and n are small.

The results also show that when $\alpha = 0.1$, although the percentage of correct identification is small, INARMA with identification produces more accurate forecasts compared to the case where the order is known. This means that, in this case, using an INARMA(0,0) forecast based on the average of the previous observations produces better results than estimating α and λ and forecasting using an INAR(1) model.

However, when the autoregressive parameter is high ($\alpha = 0.9$) and the number of observations is small ($n = 24$), although the percentage of correct identification is considerable, the difference between INARMA without identification and with identification for the Ljung-Box method is huge. This is expected because here a time series with high autocorrelation is wrongly identified as a series with no autocorrelation. Therefore, instead of putting a high weight on the last observation, the forecast is based on the INARMA(0,0) model which uses the simple average of all previous observations with equal weights. However, as the length of history increases, the percentage of correct identification also increases and the forecast accuracy of the two cases become very close. This is also true when comparing the Ljung-Box and AIC identification methods. For high values of α, the latter is considerably better than the former due to the higher percentage of correct identification and the fact that misidentification of an INAR(1) process with high autoregressive parameter as an INARMA(0,0) has a huge penalty.

Table 8-30 Accuracy of INAR(1) forecasts for Ljung-Box and AIC identification procedures

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Ljung-Box			**AIC**			**Ljung-Box**			**AIC**			**Ljung-Box**			**AIC**			**Ljung-Box**			**AIC**		
	ME	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**	**ME**	**MSE**	**MASE**
$\alpha = 0.1, \lambda = 0.5$	-0.0105	0.6244	1.0977	0.0006	0.6343	1.1239	0.0064	0.6235	1.0925	0.0104	0.6038	1.0563	-0.0056	0.5928	1.0145	0.0084	0.6196	1.0273	-0.0024	0.5691	0.9758	-0.0038	0.5660	0.9690
$\alpha = 0.5, \lambda = 0.5$	0.0047	1.0475	1.4131	0.0144	1.0372	1.3421	-0.0072	1.0201	1.2497	0.0102	0.9106	1.1900	0.0016	0.9858	1.2016	0.0023	0.8703	1.1342	0.0062	0.8633	1.1078	0.0039	0.7970	1.0655
$\alpha = 0.9, \lambda = 0.5$	0.0547	2.7057	1.9921	0.0082	1.4192	1.4380	0.0188	2.0866	1.6381	-0.0122	1.1614	1.3010	-0.0193	1.5043	1.3871	-0.0223	1.0431	1.1720	-0.0058	1.0078	1.1428	-0.0091	1.0118	1.1512
$\alpha = 0.1, \lambda = 1$	-0.0022	1.2865	0.9486	0.0063	1.3055	0.8974	-0.0019	1.1974	0.8462	-0.0011	1.2412	0.8653	-0.0053	1.1850	0.8292	-0.0045	1.1771	0.8355	0.0106	1.1682	0.8102	-0.0064	1.1372	0.7995
$\alpha = 0.5, \lambda = 1$	0.0108	2.1835	1.1859	0.0187	1.9727	1.1119	0.0284	2.1543	1.1541	-0.0017	1.8151	1.0594	0.0131	1.9543	1.0778	0.0031	1.7068	1.0146	0.0040	1.6895	0.9885	0.0046	1.5869	0.9646
$\alpha = 0.9, \lambda = 1$	0.0246	5.3777	1.7945	-0.0147	2.8294	1.3379	0.0015	4.2493	1.5135	-0.0247	2.3516	1.1871	-0.0190	2.8263	1.2406	-0.0047	2.1486	1.1096	-0.0084	2.0473	1.0789	-0.0035	2.0222	1.0807
$\alpha = 0.1, \lambda = 3$	0.0226	3.9401	0.8599	0.0091	3.8786	0.8505	-0.0248	3.6785	0.8039	0.0103	3.8087	0.8416	0.0014	3.6089	0.8094	0.0014	3.6078	0.8150	0.0065	3.4860	0.7868	-0.0025	3.4883	0.7823
$\alpha = 0.5, \lambda = 3$	0.0856	6.5598	1.1244	-0.0293	5.9107	1.0177	-0.0051	6.2115	1.0573	0.0050	5.4093	0.9960	0.0089	5.9927	1.0129	0.0133	5.1793	0.9548	0.0333	5.2292	0.9589	-0.0359	4.6773	0.9000
$\alpha = 0.9, \lambda = 3$	0.1048	15.6396	1.6548	0.0026	8.1910	1.2077	-0.0602	11.6692	1.3805	0.0020	6.9766	1.1231	0.0335	9.4088	1.2402	-0.0010	6.5496	1.0791	0.0076	6.1685	1.0349	-0.0080	6.0951	1.0283
$\alpha = 0.1, \lambda = 5$	-0.0021	6.6522	0.8473	0.0089	6.4835	0.8336	0.0023	6.1105	0.8153	0.0198	6.2402	0.8273	-0.0329	5.8728	0.7829	0.0073	6.0014	0.7918	-0.0209	5.7109	0.7718	0.0015	5.8188	0.7695
$\alpha = 0.5, \lambda = 5$	0.0828	10.9849	1.1122	-0.0084	9.6383	1.0246	0.0380	10.2881	1.0358	-0.0419	8.7787	0.9678	0.0139	9.9559	1.0261	0.0012	8.4644	0.9364	-0.0196	8.4499	0.9316	-0.0009	7.9078	0.8969
$\alpha = 0.9, \lambda = 5$	-0.0928	26.7365	1.6375	-0.0277	13.8802	1.2078	-0.0658	21.3689	1.4308	-0.0124	11.6831	1.1051	-0.0208	15.1822	1.2121	0.0341	10.8485	1.0563	0.0020	10.2432	1.0281	-0.0286	10.1068	1.0197

Table 8-31 Accuracy of INAR(1) forecasts when the order is known

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \lambda = 0.5$	0.0026	0.6614	1.1418	-0.0089	0.6286	1.0555	0.0052	0.6047	1.0237	0.0045	0.5822	0.9822
$\alpha = 0.5, \lambda = 0.5$	-0.0026	0.9466	1.3290	-0.0089	0.8743	1.1807	0.0044	0.8532	1.1570	-0.0054	0.7834	1.0534
$\alpha = 0.9, \lambda = 0.5$	0.0017	1.2082	1.3901	0.0042	1.1489	1.2894	-0.0043	1.0811	1.2176	-0.0033	1.0164	1.1486
$\alpha = 0.1, \lambda = 1$	-0.0149	1.2880	0.9376	-0.0059	1.2072	0.8496	0.0052	1.1967	0.8382	0.0029	1.1663	0.8057
$\alpha = 0.5, \lambda = 1$	-0.0195	1.8846	1.0837	-0.0018	1.7244	1.0202	-0.0045	1.6745	1.0059	0.0026	1.5905	0.9618
$\alpha = 0.9, \lambda = 1$	-0.0121	2.5216	1.2401	0.0054	2.2655	1.1594	-0.0146	2.1925	1.1262	-0.0105	2.0319	1.0686
$\alpha = 0.1, \lambda = 3$	0.0100	3.9224	0.8544	-0.0176	3.6703	0.8228	-0.0010	3.6641	0.8136	0.0085	3.4701	0.7846
$\alpha = 0.5, \lambda = 3$	-0.0308	5.6926	1.0126	-0.0035	5.1557	0.9491	-0.0078	4.9742	0.9331	0.0143	4.7749	0.9160
$\alpha = 0.9, \lambda = 3$	-0.0906	7.5862	1.1509	-0.0298	6.7494	1.1026	-0.0243	6.4376	1.0658	-0.0093	6.0230	1.0205
$\alpha = 0.1, \lambda = 5$	0.0371	6.6926	0.8565	0.0209	6.1869	0.8143	0.0282	6.0095	0.7973	0.0139	5.7529	0.7742
$\alpha = 0.5, \lambda = 5$	0.0123	9.3467	0.9840	0.0082	8.6581	0.9653	0.0116	8.3583	0.9392	-0.0034	7.8699	0.9011
$\alpha = 0.9, \lambda = 5$	-0.0013	11.9986	1.1483	-0.0126	11.2051	1.0914	-0.0467	10.8985	1.0720	0.0185	10.1102	1.0272

8.6.2.2 All-INAR(1)

In this section, it is again assumed that data is produced by either an INARMA(0,0) or an INAR(1) process. Instead of identification, we assume that the most general model, INAR(1) in this case, is used for estimation and forecasting. It is expected that if the data is in fact an INARMA(0,0) process, the estimated autoregressive parameter should be close to zero and the results confirm this. In general, the forecasting accuracy deteriorates slightly in this case compared to the case of using Ljung-Box for identification among two possible models. The results for all points in time are shown in Table 8-32.

This shows that, instead of adding an extra step to the INARMA forecasting procedure for identification, treating everything as an INAR(1) process produces close results to those with identification and it has the advantage of being simple. The degree of deterioration caused by skipping identification is on average 2 percent for both MSE and MASE.

Table 8-32 Accuracy of forecasts with identification and all-INAR(1) for INARMA(0,0) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Ljung-Box			All-INAR(1)			Ljung-Box			All-INAR(1)			Ljung-Box			All-INAR(1)			Ljung-Box			All-INAR(1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\lambda = 0.3$	0.0000	0.3509	1.1781	-0.0040	0.3610	1.1919	0.0005	0.3309	1.1476	-0.0008	0.3383	1.1552	-0.0032	0.3161	1.0892	-0.0066	0.3295	1.0741	-0.0039	0.3075	1.0115	0.0025	0.3131	1.0161
$\lambda = 0.5$	-0.0021	0.5796	1.0779	-0.0054	0.5967	1.0959	-0.0012	0.5512	1.0077	-0.0019	0.5613	1.0133	0.0009	0.5368	0.9725	0.0013	0.5561	0.9737	-0.0039	0.5183	0.9365	-0.0017	0.5191	0.9412
$\lambda = 0.7$	0.0003	0.8039	0.9807	-0.0062	0.8331	0.9786	-0.0039	0.7708	0.9005	-0.0017	0.7855	0.9152	0.0066	0.7572	0.8817	-0.0016	0.7510	0.8937	0.0048	0.7241	0.8537	0.0012	0.7374	0.8466
$\lambda = 1$	0.0012	1.1651	0.8716	-0.0028	1.1930	0.8947	0.0002	1.0979	0.8120	-0.0011	1.1178	0.8275	0.0000	1.0774	0.8025	-0.0068	1.0679	0.7956	0.0115	1.0490	0.7596	-0.0023	1.0481	0.7666
$\lambda = 3$	0.0210	3.4439	0.8065	-0.0273	3.5655	0.8128	-0.0154	3.3509	0.7703	0.0024	3.3711	0.7799	0.0087	3.2019	0.7589	-0.0083	3.2774	0.7651	0.0054	3.1061	0.7422	-0.0086	3.0890	0.7322
$\lambda = 5$	0.0099	5.8831	0.8153	-0.0158	5.8972	0.8071	-0.0067	5.5456	0.7624	-0.0150	5.4933	0.7567	0.0001	5.4289	0.7581	0.0010	5.4727	0.7642	0.0123	5.2099	0.7375	-0.0003	5.2125	0.7325
$\lambda = 20$	-0.0248	22.8486	0.7730	-0.0053	23.8074	0.8088	-0.0543	22.0121	0.7581	-0.0683	22.0701	0.7631	0.0183	21.2026	0.7420	-0.0101	22.2400	0.7589	0.0079	20.5646	0.7248	-0.0335	20.8631	0.7307

8.6.2.3 Identification among Four Processes

In this section it is assumed that data can be produced by one of the four processes that we focus on in this study: INARMA(0,0), INAR(1), INMA(1), or INARMA(1,1). As previously mentioned, two methods of identification are used. The two-stage method is based on first using the Ljung-Box Q^*-statistic of Equation 8-15 to distinguish between the INARMA(0,0) or random Poisson process from the other possible INARMA models. Then, the other three models (INAR(1), INMA(1), and INARMA(1,1)) are distinguished using the Akaike information criterion based on the expression for ARMA models:

$$\text{AIC} \approx N\text{log}\sigma_a^2 + 2m$$

Equation 8-16

As explained in section 4.5, when the sample size is small ($N/m < 40$), the following bias correction is necessary:

$$\text{AIC}_\text{C} \approx N\text{log}\sigma_a^2 + 2m + 2m(m+1)/(N-m-1)$$

Equation 8-17

On the other hand, the one-stage method only uses the AIC to select the appropriate model. As discussed in chapter 4, although the above equations have been developed for ARMA models with a Normal distribution, as the likelihood function for INMA processes has not been established in the literature and AIC is a method of identification that can be automated, we use these equations to test the performance of AIC for INARMA processes.

The percentage of correct identification for each of the four INARMA processes for both two-stage and one-stage methods is shown in Table 8-33, Table 8-34, Table 8-35, and Table 8-36.

The results of Table 8-33 confirm the results of section 8.6.2.1 in that the two-stage method identifies the INARMA(0,0) processes more frequently than the one-stage method.

Table 8-33 The percentage of correct identification for INARMA(0,0) series

Parameters	Two-stage identification				One-stage identification			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\lambda = 0.3$	94.62	95.87	96.30	94.90	89.40	89.00	88.60	80.80
$\lambda = 0.5$	93.29	94.45	94.40	95.10	91.90	88.10	84.30	73.30
$\lambda = 0.7$	92.81	94.07	95.30	95.90	88.80	82.80	78.80	69.00
$\lambda = 1$	93.54	94.32	94.10	95.10	90.40	79.60	76.00	67.20
$\lambda = 3$	93.40	92.40	94.40	94.70	82.70	72.50	65.40	57.00
$\lambda = 5$	92.90	93.40	92.80	93.50	78.50	70.90	61.70	55.00
$\lambda = 20$	92.70	93.40	94.10	95.30	76.20	66.00	61.00	51.40

For an INAR(1) case, the results of Table 8-34 confirm that when the autoregressive parameter is low, the process is misidentified in most cases for both identification methods. However, the one-stage method produces better results. On the other hand, with a high autoregressive parameter, the performance of both methods improves. The results also show that when more observations are available, the percentage of correct identification increases for both methods. The one-stage method outperforms the two-stage method in most of the cases, especially for small samples.

Table 8-34 The percentage of correct identification for INAR(1) series

Parameters	Two-stage identification				One-stage identification			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \lambda = 0.5$	2.30	2.50	2.60	2.90	9.60	15.80	21.20	26.40
$\alpha = 0.5, \lambda = 0.5$	9.50	19.80	25.80	52.70	38.40	55.60	67.70	75.80
$\alpha = 0.9, \lambda = 0.5$	30.60	55.00	61.50	73.80	73.30	71.70	71.70	72.50
$\alpha = 0.1, \lambda = 1$	1.70	1.50	1.80	3.00	13.20	19.20	22.20	30.00
$\alpha = 0.5, \lambda = 1$	10.90	18.70	26.60	49.70	46.90	59.20	65.90	72.40
$\alpha = 0.9, \lambda = 1$	28.60	52.40	62.50	71.50	74.70	71.80	72.20	69.30
$\alpha = 0.1, \lambda = 3$	1.90	2.50	1.70	2.90	16.30	22.10	22.40	28.00
$\alpha = 0.5, \lambda = 3$	11.20	16.10	24.30	50.20	47.40	57.90	60.70	68.00
$\alpha = 0.9, \lambda = 3$	24.40	46.30	63.90	75.30	75.50	74.90	72.30	75.50
$\alpha = 0.1, \lambda = 5$	2.70	2.40	2.70	2.50	17.50	22.70	22.50	26.10
$\alpha = 0.5, \lambda = 5$	12.10	17.60	26.30	44.70	50.20	59.10	61.50	68.30
$\alpha = 0.9, \lambda = 5$	27.30	49.40	63.00	73.80	76.20	75.40	73.30	73.90

As can be seen from Table 8-35, an INMA(1) process is misidentified in most of the cases regardless of the size of the moving average parameter. However, the results show that it does not affect the forecasting accuracy to a great extent. This can be seen by comparing the results of Table 8-38 and Table 8-41. The one-stage identification method again outperforms the two-stage method.

Table 8-35 The percentage of correct identification for INMA(1) series

Parameters	Two-stage identification				One-stage identification			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\beta = 0.1, \lambda = 0.5$	3.30	2.70	3.30	3.10	4.10	6.70	8.10	12.00
$\beta = 0.5, \lambda = 0.5$	4.30	4.00	4.60	7.60	13.40	19.80	23.00	23.20
$\beta = 0.9, \lambda = 0.5$	7.90	10.60	11.60	32.50	21.70	33.70	37.80	45.10
$\beta = 0.1, \lambda = 1$	4.40	3.90	3.30	1.50	5.60	6.80	8.10	11.50
$\beta = 0.5, \lambda = 1$	3.20	3.90	3.30	4.10	9.80	14.60	17.20	16.80
$\beta = 0.9, \lambda = 1$	7.80	7.60	6.70	14.90	19.80	23.50	24.70	21.50
$\beta = 0.1, \lambda = 3$	3.80	3.40	2.50	1.20	5.40	6.70	9.00	14.00
$\beta = 0.5, \lambda = 3$	2.00	2.70	2.50	2.10	10.90	14.60	15.10	11.50
$\beta = 0.9, \lambda = 3$	4.60	4.00	4.30	7.10	18.30	16.70	14.40	11.00
$\beta = 0.1, \lambda = 5$	4.80	2.30	2.50	2.00	5.70	7.40	12.10	12.10
$\beta = 0.5, \lambda = 5$	3.90	3.20	2.20	1.80	10.60	14.20	11.20	9.60
$\beta = 0.9, \lambda = 5$	3.80	4.30	3.20	5.00	17.30	15.80	13.40	7.50

The results of Table 8-36 suggest that, as expected, when the autoregressive parameter is high, the correct model is identified more often than the case with low autoregressive parameter. The identification performance improves when the length of history increases. The one-stage identification method produces better results than the two-stage method in most of the cases.

Table 8-36 The percentage of correct identification for INARMA(1,1) series

Parameters	Two-stage identification				One-stage identification			
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 24$	$n = 36$	$n = 48$	$n = 96$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.60	0.40	0.90	1.70	1.30	2.40	4.40	13.20
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.70	2.70	6.90	14.00	5.80	11.20	10.80	12.00
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	3.10	6.20	10.80	14.50	7.80	12.10	13.90	16.90
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	6.20	15.50	19.60	29.50	14.20	23.70	26.10	25.60
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.10	0.80	1.30	1.70	1.70	6.90	11.80	20.10
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	2.30	4.20	7.10	11.20	7.90	14.40	15.30	12.80
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	3.00	8.60	11.80	15.20	11.50	18.80	20.10	14.70
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	7.30	20.00	29.70	31.30	24.20	31.60	34.70	34.40
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.50	1.70	2.20	3.70	8.80	18.70	23.70	33.10
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	7.00	13.20	17.20	23.80	29.50	40.40	42.30	27.20
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	12.20	18.80	32.50	37.90	35.60	46.30	50.40	43.40
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	14.20	31.40	42.10	47.80	37.10	49.50	48.70	47.80

The accuracy of INAR(1), INMA(1) and INARMA(1,1) forecasts using ME, MSE and MASE for two identification methods are presented in Table 8-37, Table 8-38, and Table 8-39. Similar results for the cases that the order of the model is known are provided in Table 8-40, Table 8-41, and Table 8-42 for comparison reasons.

The results show that for processes with an AR component, when the autoregressive parameter is high, misidentification has a great effect on the accuracy of forecasts. However, because the one-stage identification method identifies the correct model more frequently than the two-stage method, the forecasts are closer to the case of known order. But when the autoregressive parameter is small, the effect of misidentification on forecasting accuracy is also small. For INARMA processes without an AR component, the effect of misidentification on forecasting accuracy is small, regardless of the size of the MA parameter. When the number of observations increases, forecasts with identification and without identification have similar accuracy.

Table 8-37 Accuracy of INAR(1) forecasts for one-stage and two-stage identification procedures

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \lambda = 0.5$	-0.0013	0.6357	1.1459	-0.0005	0.6347	1.1208	0.0073	0.6327	1.1044	0.0115	0.6260	1.1063	0.0008	0.6026	1.0373	0.0056	0.6005	1.0295	0.0019	0.5765	0.9804	0.0010	0.5826	0.9799
$\alpha = 0.5, \lambda = 0.5$	0.0034	1.1192	1.4704	0.0381	1.0614	1.3967	0.0278	1.0414	1.2999	0.0050	0.9311	1.2226	0.0179	1.0165	1.2303	0.0064	0.8655	1.1425	0.0200	0.8786	1.1276	-0.0055	0.8014	1.0694
$\alpha = 0.9, \lambda = 0.5$	0.0283	2.6093	1.9312	0.0279	1.4243	1.4776	-0.0005	2.0101	1.6350	0.0034	1.1610	1.2703	0.0090	1.6472	1.4275	-0.0103	1.0725	1.2157	0.0025	1.0257	1.1445	-0.0120	0.9981	1.1395
$\alpha = 0.1, \lambda = 1$	0.0188	1.3041	0.9162	0.0063	1.3144	0.9415	-0.008	1.2309	0.8614	-0.0024	1.2606	0.8758	0.0029	1.2061	0.8395	0.0071	1.2258	0.8523	0.0005	1.1699	0.8088	0.0044	1.1590	0.8104
$\alpha = 0.5, \lambda = 1$	0.0409	2.3062	1.2365	0.0196	2.0526	1.1509	0.0131	2.0748	1.1167	0.0117	1.8415	1.0550	0.0019	1.9812	1.0876	0.0093	1.7273	1.0262	0.0109	1.7521	1.0034	0.0039	1.6017	0.9651
$\alpha = 0.9, \lambda = 1$	-0.018	5.2825	1.7681	-0.0212	2.7024	1.2864	0.0258	4.2144	1.4907	0.0037	2.2686	1.1378	0.0222	3.1025	1.2932	-0.0043	2.1714	1.1203	0.0087	2.0297	1.0845	-0.0012	2.0015	1.0693
$\alpha = 0.1, \lambda = 3$	-0.0055	3.7732	0.8183	-0.0310	3.8137	0.8345	-0.0096	3.7235	0.8334	0.0042	3.7900	0.8337	-0.0036	3.5162	0.7995	-0.0041	3.6297	0.8083	0.0098	3.5198	0.7902	-0.0184	3.4845	0.7845
$\alpha = 0.5, \lambda = 3$	-0.0356	6.6692	1.0898	0.0256	6.0453	1.0472	0.0322	6.4381	1.0708	0.0170	5.5157	0.9955	0.0377	6.0613	1.0365	-0.0103	5.1944	0.9556	0.0263	5.1238	0.9517	0.0027	4.7714	0.9155
$\alpha = 0.9, \lambda = 3$	0.0466	17.2237	1.6973	-0.0198	8.0490	1.2229	0.0386	12.4798	1.4291	0.0090	6.7598	1.1070	0.0380	8.8346	1.1905	0.0301	6.4818	1.0660	0.0107	6.1671	1.0259	0.0099	6.0402	1.0160
$\alpha = 0.1, \lambda = 5$	-0.0215	6.3926	0.8466	0.0440	6.7467	0.8632	0.0314	6.1703	0.8199	0.0022	6.1377	0.8190	-0.0108	6.0691	0.8034	-0.0241	6.1470	0.8087	0.0073	5.7416	0.7720	-0.0252	5.9063	0.7726
$\alpha = 0.5, \lambda = 5$	-0.0479	10.7311	1.0682	-0.0116	9.9783	1.0548	0.1183	10.4885	1.0480	-0.0001	9.0103	0.9843	-0.0082	9.7144	1.0139	-0.0479	8.4696	0.9420	0.0108	8.5805	0.9368	-0.0078	7.9000	0.9092
$\alpha = 0.9, \lambda = 5$	-0.0995	26.0358	1.6186	0.1313	13.5067	1.1973	0.0334	21.0114	1.4054	0.0554	11.3171	1.0809	-0.0524	15.4657	1.1925	0.0363	10.8754	1.0706	-0.0140	10.2025	1.0191	-0.0036	10.0731	1.0171

Table 8-38 Accuracy of INMA(1) forecasts for one-stage and two-stage identification procedures

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\beta = 0.1, \lambda = 0.5$	0.0263	0.6370	1.1790	0.0006	0.6474	1.1537	0.0081	0.6080	1.0762	-0.0042	0.6022	1.0458	0.0030	0.5913	1.0025	0.0005	0.5732	0.9729	0.0064	0.5719	0.9804	-0.0088	0.6604	0.7515
$\beta = 0.5, \lambda = 0.5$	0.0166	0.8479	1.3582	0.0210	0.8392	1.2970	0.0125	0.8181	1.2071	-0.0001	0.7935	1.1635	0.0022	0.7871	1.1224	0.0013	0.7284	1.0496	0.0082	0.7597	1.0724	0.0401	0.9122	0.9183
$\beta = 0.9, \lambda = 0.5$	0.0246	1.0729	1.4336	-0.0101	1.9689	1.1383	0.0207	1.0169	1.3105	0.0269	0.9813	1.2952	0.0240	1.0123	1.2547	0.0304	0.8882	1.1203	0.0250	0.9228	1.1422	0.1589	1.0747	1.0178
$\beta = 0.1, \lambda = 1$	-0.0003	1.2975	0.9130	-0.0104	1.2574	0.8783	-0.0070	1.2071	0.8535	-0.0014	1.2389	0.8692	0.0017	1.1671	0.8266	-0.0103	1.1444	0.8014	0.0036	1.1466	0.8073	-0.0335	1.2159	0.8080
$\beta = 0.5, \lambda = 1$	0.0098	1.7287	1.0880	0.0241	1.7225	1.0816	0.0153	1.6183	1.0208	-0.0065	1.5827	0.9817	-0.0051	1.6029	1.0149	-0.0020	1.4386	0.9239	0.0077	1.5257	0.9653	0.0177	1.5554	0.9547
$\beta = 0.9, \lambda = 1$	0.0081	2.1437	1.2136	-0.0131	1.9568	1.1574	0.0027	2.0207	1.1294	0.0036	1.8299	1.0827	0.0221	2.0155	1.1022	0.0179	1.6855	0.9954	0.0025	1.7263	1.0134	0.0595	1.7732	1.0167
$\beta = 0.1, \lambda = 3$	-0.0143	3.8219	0.8451	0.0170	3.9592	0.8653	0.0095	3.6174	0.8141	-0.0217	3.7610	0.8249	-0.0132	3.5934	0.8019	0.0052	3.4847	0.7859	-0.0158	3.3791	0.7756	0.0006	3.5011	0.8056
$\beta = 0.5, \lambda = 3$	0.0029	5.1085	0.9854	0.0143	5.2555	0.9867	0.0086	4.8587	0.9522	-0.0237	4.8094	0.9361	0.0259	4.7425	0.9197	-0.0040	4.2445	0.8681	0.0024	4.5513	0.8899	-0.0040	4.2567	0.8749
$\beta = 0.9, \lambda = 3$	-0.0104	6.3963	1.1064	-0.0283	6.0747	1.0769	0.0162	5.9595	1.0526	0.0190	5.4653	1.0106	0.0191	5.7408	1.0273	-0.0034	4.8673	0.9248	-0.0221	5.0070	0.9420	-0.0028	4.8710	0.9265
$\beta = 0.1, \lambda = 5$	0.0349	6.4620	0.8532	-0.0388	6.5789	0.8436	-0.0228	6.1804	0.8130	-0.0241	6.1703	0.8007	-0.0035	5.9428	0.7934	0.0168	5.7806	0.7811	0.0003	5.7474	0.7794	0.0160	5.7818	0.7843
$\beta = 0.5, \lambda = 5$	0.0000	8.3916	0.9442	-0.0841	8.5133	0.9644	0.0077	8.0429	0.9225	-0.0184	7.8372	0.9231	0.0038	8.0922	0.9232	0.0028	7.1503	0.8623	0.0291	7.5896	0.8869	0.0027	7.1503	0.8628
$\beta = 0.9, \lambda = 5$	0.0320	10.5829	1.1034	-0.0552	9.9051	1.0576	-0.0166	10.2863	1.0590	0.0005	8.9502	0.9893	-0.0596	9.3927	0.9941	0.0135	7.9759	0.9193	-0.0112	8.3874	0.9427	0.0135	7.9764	0.9194

Table 8-39 Accuracy of INARMA(1,1) forecasts for one-stage and two-stage identification procedures

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification			Two-stage identification			One-stage identification		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.0011	0.7193	1.1489	0.0064	0.7234	1.2251	0.0071	0.6856	1.0945	0.0163	0.6807	1.1061	0.0040	0.6758	1.0704	0.0071	0.6736	1.0554	0.0059	0.6447	1.0204	-0.0001	0.6425	1.0050
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.0368	1.3416	1.5159	0.0243	1.2303	1.4021	0.0201	1.2114	1.2766	0.0094	1.0992	1.2406	0.0173	1.1478	1.2195	0.0162	1.0410	1.1899	0.0159	1.0410	1.1174	0.0138	1.0175	1.1117
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.0536	1.9040	1.5660	0.0198	1.6235	1.4007	0.0186	1.6852	1.3458	0.0318	1.4292	1.2942	0.0301	1.5869	1.3092	0.0198	1.3254	1.1971	0.0064	1.1989	1.0976	0.0049	1.1844	1.1020
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.1342	2.8645	2.0436	0.0480	1.5286	1.4458	0.0377	2.1973	1.6213	0.0289	1.3400	1.3159	0.0277	1.6027	1.3899	0.0310	1.1833	1.2270	0.0105	1.1303	1.1566	0.0057	1.1247	1.1435
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.0004	1.4375	0.9427	-0.0167	1.4091	0.9407	0.0130	1.3668	0.8973	0.0079	1.3705	0.9013	-0.0038	1.3390	0.8764	0.0020	1.3314	0.8871	-0.0064	1.2803	0.8575	-0.0080	1.2693	0.8538
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.0137	2.5594	1.2132	0.0161	2.3860	1.1767	-0.0010	2.4152	1.1409	-0.0137	2.1568	1.0963	0.0109	2.3587	1.0905	0.0001	1.9950	1.0265	0.0028	2.0246	1.0093	0.0032	1.9137	0.9745
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.0028	3.7251	1.3345	-0.0434	3.0511	1.1857	-0.0317	3.4130	1.2229	0.0026	2.7933	1.1045	0.0225	3.0826	1.1397	0.0027	2.5683	1.0629	-0.0053	2.3995	1.0005	-0.0111	2.2977	0.9876
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.0911	6.0831	1.7798	-0.0259	3.0317	1.2791	0.0230	4.8164	1.5115	-0.0462	2.4860	1.1595	-0.0248	3.4115	1.2711	-0.0062	2.3640	1.0960	-0.0036	2.2186	1.0714	-0.0052	2.1915	1.0725
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	-0.0126	7.1217	0.8862	0.0116	7.3120	0.8927	0.0082	6.8366	0.8465	-0.0300	6.7345	0.8344	0.0131	6.6983	0.8327	0.0088	6.6701	0.8331	-0.0071	6.4563	0.8119	-0.0136	6.3122	0.8035
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.0523	12.8778	1.1181	0.0211	11.8098	1.0700	0.0125	11.8990	1.0497	-0.0111	10.4597	0.9831	0.0224	11.3640	1.0295	-0.0163	9.9553	0.9482	-0.0102	9.9313	0.9337	-0.0177	9.2879	0.9064
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	-0.0698	18.7993	1.2511	-0.0942	14.7010	1.0942	0.0161	16.7775	1.1609	-0.0720	12.9606	1.0033	-0.0238	15.2377	1.0874	-0.0601	12.3897	0.9876	0.0213	11.8909	0.9624	-0.0353	11.4452	0.9463
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.0827	29.4367	1.6336	0.0109	15.2016	1.2041	-0.0115	23.6024	1.4174	-0.0590	12.3696	1.0847	-0.0741	16.4158	1.2023	-0.0188	11.9975	1.0709	-0.0215	11.0569	1.0183	-0.0287	11.1985	1.0193

Table 8-40 Accuracy of INAR(1) forecasts when the order in known

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \lambda = 0.5$	0.0026	0.6614	1.1418	-0.0089	0.6286	1.0555	0.0052	0.6047	1.0237	0.0045	0.5822	0.9822
$\alpha = 0.5, \lambda = 0.5$	-0.0026	0.9466	1.3290	-0.0089	0.8743	1.1807	0.0044	0.8532	1.1570	-0.0054	0.7834	1.0534
$\alpha = 0.9, \lambda = 0.5$	0.0017	1.2082	1.3901	0.0042	1.1489	1.2894	-0.0043	1.0811	1.2176	-0.0033	1.0164	1.1486
$\alpha = 0.1, \lambda = 1$	-0.0149	1.2880	0.9376	-0.0059	1.2072	0.8496	0.0052	1.1967	0.8382	0.0029	1.1663	0.8057
$\alpha = 0.5, \lambda = 1$	-0.0195	1.8846	1.0837	-0.0018	1.7244	1.0202	-0.0045	1.6745	1.0059	0.0026	1.5905	0.9618
$\alpha = 0.9, \lambda = 1$	-0.0121	2.5216	1.2401	0.0054	2.2655	1.1594	-0.0146	2.1925	1.1262	-0.0105	2.0319	1.0686
$\alpha = 0.1, \lambda = 3$	0.0100	3.9224	0.8544	-0.0176	3.6703	0.8228	-0.0010	3.6641	0.8136	0.0085	3.4701	0.7846
$\alpha = 0.5, \lambda = 3$	-0.0308	5.6926	1.0126	-0.0035	5.1557	0.9491	-0.0078	4.9742	0.9331	0.0143	4.7749	0.9160
$\alpha = 0.9, \lambda = 3$	-0.0906	7.5862	1.1509	-0.0298	6.7494	1.1026	-0.0243	6.4376	1.0658	-0.0093	6.0230	1.0205
$\alpha = 0.1, \lambda = 5$	0.0371	6.6926	0.8565	0.0209	6.1869	0.8143	0.0282	6.0095	0.7973	0.0139	5.7529	0.7742
$\alpha = 0.5, \lambda = 5$	0.0123	9.3467	0.9840	0.0082	8.6581	0.9653	0.0116	8.3583	0.9392	-0.0034	7.8699	0.9011
$\alpha = 0.9, \lambda = 5$	-0.0013	11.9986	1.1483	-0.0126	11.2051	1.0914	-0.0467	10.8985	1.0720	0.0185	10.1102	1.0272

Table 8-41 Accuracy of INMA(1) forecasts when the order in known

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\beta = 0.1, \lambda = 0.5$	-0.0031	0.6295	1.1297	0.0027	0.6035	1.0473	0.0114	0.5993	1.0493	0.0007	0.5702	0.9736
$\beta = 0.5, \lambda = 0.5$	0.0395	0.8793	1.3407	0.0279	0.8552	1.2527	0.0229	0.7997	1.1494	0.0031	0.7885	1.0948
$\beta = 0.9, \lambda = 0.5$	0.0767	1.1019	1.5472	0.0671	1.0609	1.3632	0.0628	1.0229	1.2606	0.0445	0.9878	1.1797
$\beta = 0.1, \lambda = 1$	0.0234	1.2748	0.9163	0.0000	1.2038	0.8547	0.0082	1.1724	0.8365	0.0048	1.1313	0.7957
$\beta = 0.5, \lambda = 1$	0.0347	1.7554	1.0666	0.0440	1.6455	1.0247	0.0310	1.6074	1.0005	0.0204	1.5302	0.9648
$\beta = 0.9, \lambda = 1$	0.1249	2.2869	1.2485	0.1025	2.1650	1.1687	0.1148	2.0762	1.1223	0.1116	1.9944	1.0920
$\beta = 0.1, \lambda = 3$	0.0452	3.9039	0.8547	0.0402	3.6622	0.8158	0.0150	3.5353	0.7960	0.0062	3.4237	0.7767
$\beta = 0.5, \lambda = 3$	0.1260	5.2415	0.9971	0.0781	4.9565	0.9450	0.0929	4.8942	0.9424	0.0608	4.6821	0.9036
$\beta = 0.9, \lambda = 3$	0.2970	6.7422	1.1268	0.3209	6.5993	1.1023	0.2796	6.2599	1.0587	0.2718	6.0678	1.0364
$\beta = 0.1, \lambda = 5$	0.0838	6.4549	0.8431	0.0132	6.2194	0.8182	0.0226	5.9292	0.7937	0.0111	5.7446	0.7806
$\beta = 0.5, \lambda = 5$	0.1407	8.7278	0.9676	0.1984	8.3020	0.9417	0.1738	8.0993	0.9245	0.1269	7.8404	0.9063
$\beta = 0.9, \lambda = 5$	0.4387	11.6802	1.1864	0.4081	10.9042	1.0713	0.4804	10.7677	1.0895	0.4479	10.1863	1.0242

Table 8-42 Accuracy of INARMA(1,1) forecasts when the order in known

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	-0.0007	0.7482	1.2163	-0.0066	0.7138	1.1186	0.0019	0.6777	1.0787	-0.0047	0.6468	1.0076
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.0401	1.2112	1.4173	0.0336	1.1340	1.2974	0.0267	1.0734	1.1958	0.0163	1.0429	1.1248
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.0485	1.5791	1.4864	0.0209	1.4070	1.2666	0.0445	1.3825	1.1924	0.0261	1.2261	1.1137
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.0920	1.4003	1.4228	0.0532	1.2843	1.3142	0.0487	1.2226	1.2608	0.0324	1.1110	1.1489
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	-0.0041	1.4908	0.9689	-0.0108	1.4146	0.9145	-0.0129	1.3358	0.8925	-0.0119	1.2833	0.8575
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.0192	2.3265	1.1462	-0.0185	2.2569	1.0930	0.0368	2.2224	1.0713	0.0149	2.0900	1.0210
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.0296	3.1322	1.1914	0.0250	2.7101	1.1033	0.0409	2.6748	1.0798	0.0220	2.4363	1.0245
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.0665	2.8324	1.2617	0.0494	2.5623	1.1748	0.0640	2.4301	1.1219	0.0413	2.2588	1.0786
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.0279	7.4165	0.8824	-0.0070	6.9501	0.8502	0.0011	6.7534	0.8307	0.0006	6.4253	0.8079
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.0436	11.2147	1.0506	0.0580	10.5085	0.9941	0.0521	10.1101	0.9589	0.0606	9.7366	0.9338
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.1082	14.0754	1.0425	0.1239	13.0227	1.0240	0.1074	12.7060	0.9952	0.1065	11.8259	0.9556
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.1214	13.8720	1.1795	0.1895	12.4636	1.0932	0.1569	12.0389	1.0813	0.0906	11.2782	1.0311

8.6.2.4 All-INARMA(1,1)

In this section the method of section 8.6.2.2 is extended to include all four processes. Therefore data can be produced by either an INARMA(0,0), INAR(1), INMA(1) or an INARMA(1,1) process. Then, for estimation of parameters and forecasting, an INARMA(1,1) process is used. We expect that when data is in fact INARMA(0,0) the estimated autoregressive and moving average parameters (α, β) will be close to zero, and for INAR(1) and INMA(1) data, the estimated α or β will be close to zero, respectively.

The results for all points in time are shown in Table 8-43, Table 8-44, and Table 8-45. The results show that, for INARMA(0,0), identification produces better forecasts than the all-INARMA(1,1) approach. When the number of observations increases, the results of two approaches become close. For $n = 96$, the degree of improvement by using identification rather than all-INARMA(1,1) is on average 2.3 percent.

For INAR(1) and INMA(1) processes, when the number of observations is small, the all-INARMA(1,1) approach produces better results in many cases. But when the number of observations increases, the results of identification improve and the two methods produce close results.

Based on the results of Table 8-43, Table 8-44, and Table 8-45, using the most general model could be a good substitute for identification especially when less data is available as is often the case for intermittent demand data. Although it has not been looked at in the literature, the results suggest that treating the data as the general INARMA process and eliminating the complexity of identification, can be considered as a potentially useful approach.

The forecast accuracy of all-INARMA(1,1) is compared to those of all-INAR(1) in Appendix 8.G. It is expected that for INARMA(0,0) and INAR(1) series the latter outperforms the former and the results confirm this for most of the cases. However, the results show that, even for INMA(1) and INARMA(1,1) series, and even for high moving average parameters, all-INAR(1) method produces more accurate forecasts (in terms of MSE and MASE) than all-INARMA(1,1) method in most of the cases. The difference increases for longer history.

Table 8-43 Accuracy of forecasts with identification and all-INARMA(1,1) for INARMA(0,0) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Two-stage identification			All-INARMA(1,1)			Two-stage identification			All-INARMA(1,1)			Two-stage identification			All-INARMA(1,1)			Two-stage identification			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\lambda = 0.3$	0.0020	0.3496	1.1877	-0.0032	0.3692	1.2046	0.0019	0.3307	1.1525	-0.0048	0.3470	1.1612	0.0003	0.3190	1.0924	-0.0064	0.3374	1.1003	0.0028	0.3146	1.0229	-0.0032	0.3182	1.0161
$\lambda = 0.5$	-0.0010	0.5789	1.0799	-0.0088	0.6195	1.1038	0.0002	0.5493	1.0109	-0.0072	0.5774	1.0240	-0.0050	0.5209	0.9639	-0.0114	0.5404	0.9630	-0.0054	0.5099	0.9213	0.0011	0.5416	0.9403
$\lambda = 0.7$	0.0008	0.8074	0.9693	-0.0137	0.8690	0.9951	0.0011	0.7706	0.9111	-0.0057	0.8129	0.9260	-0.0023	0.7429	0.8764	-0.0193	0.7647	0.8797	-0.0022	0.7186	0.8378	-0.0010	0.7351	0.8524
$\lambda = 1$	-0.0001	1.1538	0.8665	-0.0203	1.2508	0.8980	0.0011	1.1076	0.8157	-0.0111	1.1551	0.8358	-0.0152	1.0705	0.7820	-0.0017	1.1163	0.8011	-0.0061	1.0289	0.7508	-0.0045	1.0540	0.7640
$\lambda = 3$	0.0042	3.4186	0.8175	-0.0324	3.5960	0.8346	0.0064	3.3024	0.7818	-0.0082	3.3893	0.7871	0.0162	3.2304	0.7654	-0.0151	3.3233	0.7691	-0.0025	3.0952	0.7386	-0.0081	3.1778	0.7455
$\lambda = 5$	0.0161	5.8873	0.8155	-0.0507	6.0384	0.8235	-0.0107	5.5208	0.7711	-0.0144	5.7687	0.7783	0.0042	5.4208	0.7613	0.0135	5.5495	0.7689	-0.0153	5.2324	0.7339	-0.0056	5.1751	0.7306
$\lambda = 20$	0.0509	22.9286	0.7872	-0.0146	24.9564	0.8225	-0.0421	21.9667	0.7702	-0.0514	22.8435	0.7773	-0.0156	21.5521	0.7400	-0.0269	22.0828	0.7551	-0.0252	20.5910	0.7322	-0.0400	21.1039	0.7334

Table 8-44 Accuracy of forecasts with identification and all-INARMA(1,1) for INAR(1) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \lambda = 0.5$	-0.0005	0.6347	1.1208	-0.0008	0.6699	1.1275	0.0115	0.6260	1.1063	0.0049	0.6447	1.0710	0.0056	0.6005	1.0295	0.0111	0.6389	1.0340	0.0010	0.5826	0.9799	0.0010	0.5857	0.9845
$\alpha = 0.5, \lambda = 0.5$	0.0381	1.0614	1.3967	0.0130	1.0056	1.3278	0.0050	0.9311	1.2226	0.0039	0.9436	1.2602	0.0064	0.8655	1.1425	0.0041	0.9131	1.1864	-0.0055	0.8014	1.0694	-0.0033	0.8186	1.0727
$\alpha = 0.9, \lambda = 0.5$	0.0279	1.4243	1.4776	0.0347	1.2681	1.4405	0.0034	1.1610	1.2703	0.0344	1.1681	1.3181	-0.0103	1.0725	1.2157	0.0066	1.0919	1.2150	-0.0120	0.9981	1.1395	0.0206	1.0235	1.1707
$\alpha = 0.1, \lambda = 1$	0.0063	1.3144	0.9415	-0.0152	1.3755	0.9455	-0.0024	1.2606	0.8758	-0.0028	1.2778	0.8819	0.0071	1.2258	0.8523	-0.0066	1.1947	0.8427	0.0044	1.1590	0.8104	0.0045	1.1875	0.8235
$\alpha = 0.5, \lambda = 1$	0.0196	2.0526	1.1509	-0.0092	1.9908	1.1188	0.0117	1.8415	1.0550	0.0050	1.8491	1.0640	0.0093	1.7273	1.0262	-0.0046	1.7735	1.0251	0.0039	1.6017	0.9651	0.0052	1.6469	0.9944
$\alpha = 0.9, \lambda = 1$	-0.0212	2.7024	1.2864	0.0544	2.4325	1.2304	0.0037	2.2686	1.1378	0.0378	2.3301	1.1748	-0.0043	2.1714	1.1203	0.0484	2.1853	1.1423	-0.0012	2.0015	1.0693	0.0312	2.0343	1.0747
$\alpha = 0.1, \lambda = 3$	-0.0310	3.8137	0.8345	-0.0275	4.1292	0.8818	0.0042	3.7900	0.8337	-0.0177	3.7457	0.8236	-0.0041	3.6297	0.8083	-0.0144	3.6442	0.8115	-0.0184	3.4845	0.7845	-0.0222	3.4929	0.7828
$\alpha = 0.5, \lambda = 3$	0.0256	6.0453	1.0472	0.0237	5.8020	1.0279	0.0170	5.5157	0.9955	-0.0291	5.4831	0.9848	-0.0103	5.1944	0.9556	-0.0027	5.2166	0.9710	0.0027	4.7714	0.9155	0.0268	4.9748	0.9319
$\alpha = 0.9, \lambda = 3$	-0.0198	8.0490	1.2229	0.1097	7.3168	1.1422	0.0090	6.7598	1.1070	0.1077	6.9143	1.1046	0.0301	6.4818	1.0660	0.0627	6.5204	1.0742	0.0099	6.0402	1.0160	0.0612	6.1440	1.0290
$\alpha = 0.1, \lambda = 5$	0.0440	6.7467	0.8632	0.0124	6.5975	0.8522	0.0022	6.1377	0.8190	0.0016	6.3370	0.8206	-0.0241	6.1470	0.8087	-0.0078	6.2455	0.8197	-0.0252	5.9063	0.7726	-0.0063	5.8121	0.7805
$\alpha = 0.5, \lambda = 5$	-0.0116	9.9783	1.0548	0.0798	9.5203	1.0114	-0.0001	9.0103	0.9843	0.0050	8.8544	0.9603	-0.0479	8.4696	0.9420	0.0325	8.4735	0.9460	-0.0078	7.9000	0.9092	-0.0003	7.9638	0.9095
$\alpha = 0.9, \lambda = 5$	0.1313	13.5067	1.1973	0.0773	12.2386	1.1385	0.0554	11.3171	1.0809	0.1196	11.3739	1.0820	0.0363	10.8754	1.0706	0.1021	10.9289	1.0679	-0.0036	10.0731	1.0171	0.0956	10.1452	1.0232

Table 8-45 Accuracy of forecasts with identification and all-INARMA(1,1) for INMA(1) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)			One-stage identification			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\beta = 0.1, \lambda = 0.5$	0.0006	0.6474	1.1537	-0.0056	0.6500	1.1831	-0.0042	0.6022	1.0458	-0.0038	0.6155	1.0589	0.0005	0.5732	0.9729	-0.0071	0.6070	1.0025	-0.0088	0.6604	0.7515	0.0063	0.5733	0.9809
$\beta = 0.5, \lambda = 0.5$	0.0210	0.8392	1.2970	0.0205	0.8681	1.3428	-0.0001	0.7935	1.1635	-0.0151	0.8047	1.1628	0.0013	0.7284	1.0496	0.0011	0.7815	1.1069	0.0401	0.9122	0.9183	-0.0038	0.7423	1.0637
$\beta = 0.9, \lambda = 0.5$	-0.0101	1.9689	1.1383	0.0430	1.0089	1.3941	0.0269	0.9813	1.2952	0.0352	0.9383	1.2649	0.0304	0.8882	1.1203	0.0122	0.9436	1.1917	0.1589	1.0747	1.0178	0.0250	0.9208	1.1448
$\beta = 0.1, \lambda = 1$	-0.0104	1.2574	0.8783	-0.0202	1.3447	0.9356	-0.0014	1.2389	0.8692	-0.0154	1.2458	0.8487	-0.0103	1.1444	0.8014	-0.0017	1.1956	0.8443	-0.0335	1.2159	0.8080	-0.0016	1.1525	0.8117
$\beta = 0.5, \lambda = 1$	0.0241	1.7225	1.0816	0.0060	1.6743	1.0374	-0.0065	1.5827	0.9817	-0.0128	1.5905	0.9873	-0.0020	1.4386	0.9239	0.0030	1.5879	0.9894	0.0177	1.5554	0.9547	-0.0077	1.4733	0.9413
$\beta = 0.9, \lambda = 1$	-0.0131	1.9568	1.1574	0.0145	1.9996	1.1350	0.0036	1.8299	1.0827	0.0180	1.9324	1.0960	0.0179	1.6855	0.9954	0.0165	1.8242	1.0545	0.0595	1.7732	1.0167	0.0184	1.7825	1.0406
$\beta = 0.1, \lambda = 3$	0.0170	3.9592	0.8653	-0.0153	3.8953	0.8642	-0.0217	3.7610	0.8249	0.0022	3.7206	0.8211	0.0052	3.4847	0.7859	0.0041	3.6216	0.7983	0.0006	3.5011	0.8056	-0.0215	3.5003	0.7873
$\beta = 0.5, \lambda = 3$	0.0143	5.2555	0.9867	-0.0306	4.8945	0.9557	-0.0237	4.8094	0.9361	-0.0137	4.7351	0.9278	-0.0040	4.2445	0.8681	-0.0013	4.6193	0.9143	-0.0040	4.2567	0.8749	0.0061	4.3870	0.8783
$\beta = 0.9, \lambda = 3$	-0.0283	6.0747	1.0769	0.0763	5.8039	1.0545	0.0190	5.4653	1.0106	0.0192	5.5915	1.0116	-0.0034	4.8673	0.9248	0.0118	5.2638	0.9726	-0.0028	4.8710	0.9265	0.0439	5.1879	0.9612
$\beta = 0.1, \lambda = 5$	-0.0388	6.5789	0.8436	0.0108	6.5867	0.8570	-0.0241	6.1703	0.8007	-0.0108	6.2327	0.8241	0.0168	5.7806	0.7811	-0.0045	6.0357	0.7906	0.0160	5.7818	0.7843	-0.0103	5.7236	0.7712
$\beta = 0.5, \lambda = 5$	-0.0841	8.5133	0.9644	-0.0335	8.3565	0.9718	-0.0184	7.8372	0.9231	0.0231	7.7461	0.9228	0.0028	7.1503	0.8623	0.0136	7.6924	0.8954	0.0027	7.1503	0.8628	0.0160	7.2509	0.8698
$\beta = 0.9, \lambda = 5$	-0.0552	9.9051	1.0576	0.0927	9.8826	1.0531	0.0005	8.9502	0.9893	0.0643	8.8871	0.9798	0.0135	7.9759	0.9193	0.0683	8.9059	0.9673	0.0135	7.9764	0.9194	0.0524	8.4545	0.9443

8.6.2.5 All-INAR(1) vs Benchmark Methods

Based on the argument in the previous section, the degree of improvement by treating all INARMA series as an INAR(1) model over the benchmark methods is investigated in this section. The MSE results for INARMA(0,0), INMA(1), and INARMA(1,1) series are shown in Table 8-46 to Table 8-49. The results for INAR(1) series are the same as the results of known order (Table 8-18 and Table 8-19). The results are for the case that all points in time are considered.

As previously mentioned in section 8.6.1, there was a slight improvement over the benchmark methods when demand is INARMA(0,0) or INMA(1). This was for the case that the order of the INARMA model was known. Considering the fact that the identification errors result in deterioration of forecasting accuracy for INARMA models, we except benchmark methods to outperform INARMA especially for more sparse demand. The results of Table 8-46 and Table 8-47 confirm this.

Table 8-46 $MSE_{INARMA} / MSE_{Benchmark}$ for INARMA(0,0) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.8998	0.9435	0.9473	0.9746	0.9991	1.0012	1.0070	1.0230	1.0239	0.9918	0.9997	1.0000
$\lambda = 0.5$	0.9838	1.0119	1.0136	0.9998	1.0150	1.0159	1.0045	1.0157	1.0163	0.9756	0.9850	0.9854
$\lambda = 0.7$	1.0048	1.0262	1.0280	0.9922	1.0049	1.0054	0.9845	0.9952	0.9955	0.9663	0.9764	0.9767
$\lambda = 1$	1.0031	1.0179	1.0181	0.9785	0.9887	0.9885	0.9440	0.9533	0.9531	0.9624	0.9719	0.9716
$\lambda = 3$	1.0170	1.0095	1.0051	0.9479	0.9424	0.9385	0.9496	0.9428	0.9388	0.9198	0.9126	0.9087
$\lambda = 5$	0.9647	0.9399	0.9321	0.9383	0.9128	0.9050	0.9593	0.9364	0.9288	0.9290	0.9048	0.8972
$\lambda = 20$	0.9677	0.8339	0.8063	0.9323	0.8030	0.7762	0.9686	0.8339	0.8058	0.9276	0.7982	0.7712

Table 8-47 $MSE_{INARMA} / MSE_{Benchmark}$ for INMA(1) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	1.0350	1.0693	1.0717	0.9922	1.0109	1.0122	0.9695	0.9813	0.9818	0.9558	0.9678	0.9683
$\beta = 0.5, \lambda = 0.5$	0.8597	0.8923	0.8941	0.8599	0.8828	0.8842	0.8933	0.9129	0.9141	0.8192	0.8369	0.8380
$\beta = 0.9, \lambda = 0.5$	0.7818	0.8177	0.8200	0.7585	0.7812	0.7826	0.7387	0.7571	0.7580	0.7201	0.7387	0.7398
$\beta = 0.1, \lambda = 1$	1.0022	1.0157	1.0157	0.9694	0.9819	0.9818	0.9695	0.9804	0.9801	0.9598	0.9705	0.9703
$\beta = 0.5, \lambda = 1$	0.9325	0.9471	0.9468	0.9167	0.9269	0.9261	0.8776	0.8891	0.8885	0.8698	0.8809	0.8803
$\beta = 0.9, \lambda = 1$	0.8373	0.8436	0.8421	0.7809	0.7884	0.7873	0.7796	0.7868	0.7856	0.7468	0.7528	0.7515
$\beta = 0.1, \lambda = 3$	0.9677	0.9569	0.9521	0.9527	0.9430	0.9383	0.9385	0.9276	0.9228	0.9381	0.9295	0.9250
$\beta = 0.5, \lambda = 3$	0.9425	0.9215	0.9143	0.9388	0.9173	0.9101	0.9063	0.8860	0.8791	0.8884	0.8678	0.8610
$\beta = 0.9, \lambda = 3$	0.8611	0.8259	0.8166	0.8248	0.7942	0.7855	0.8116	0.7835	0.7752	0.8002	0.7705	0.7621
$\beta = 0.1, \lambda = 5$	0.9625	0.9316	0.9227	0.9444	0.9198	0.9116	0.9639	0.9356	0.9269	0.9229	0.8952	0.8868
$\beta = 0.5, \lambda = 5$	0.9521	0.9031	0.8903	0.9356	0.8843	0.8713	0.9162	0.8651	0.8525	0.8655	0.8224	0.8108
$\beta = 0.9, \lambda = 5$	0.8754	0.8149	0.8001	0.8617	0.8037	0.7889	0.8093	0.7546	0.7409	0.7908	0.7355	0.7219

The results for INARMA(1,1) series again show that with the presence of a high autoregressive parameter, INARMA has a considerably smaller MSE than the benchmark methods. The results of MASE also confirm this (see Appendix 8.H).

Table 8-48 $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.2 for INARMA(1,1) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9452	0.9745	0.9759	0.9301	0.9500	0.9513	0.9733	0.9899	0.9910	0.9297	0.9444	0.9451
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.7664	0.7989	0.8007	0.7325	0.7564	0.7581	0.7616	0.7811	0.7821	0.7025	0.7196	0.7206
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.6855	0.7039	0.7039	0.6828	0.6985	0.6990	0.6243	0.6358	0.6359	0.6193	0.6332	0.6336
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.5274	0.4433	0.4294	0.5164	0.4542	0.4414	0.4968	0.4385	0.4264	0.4740	0.4208	0.4091
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9771	0.9922	0.9922	0.9486	0.9621	0.9620	0.9571	0.9695	0.9693	0.9168	0.9284	0.9282
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.8802	0.8867	0.8851	0.7635	0.7730	0.7721	0.7569	0.7619	0.7605	0.7493	0.7556	0.7544
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.7720	0.7663	0.7629	0.7394	0.7370	0.7340	0.6899	0.6858	0.6828	0.6710	0.6701	0.6675
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.5481	0.4206	0.3994	0.5058	0.3942	0.3750	0.4906	0.3856	0.3671	0.4525	0.3597	0.3428
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9920	0.9531	0.9426	0.9574	0.9233	0.9135	0.9436	0.9097	0.8999	0.9182	0.8846	0.8750
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9232	0.8501	0.8332	0.8544	0.7861	0.7705	0.8255	0.7633	0.7485	0.7985	0.7393	0.7249
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8379	0.7406	0.7199	0.7689	0.6783	0.6591	0.7251	0.6455	0.6280	0.7068	0.6278	0.6106
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.5482	0.2488	0.2197	0.5407	0.2289	0.2017	0.4987	0.2170	0.1916	0.4535	0.2013	0.1779

Table 8-49 $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.5 for INARMA(1,1) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.8646	0.9584	0.9809	0.8208	0.9087	0.9414	0.8490	0.9381	0.9726	0.8068	0.8896	0.9256
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.7087	0.7961	0.7972	0.6459	0.7323	0.7446	0.6773	0.7553	0.7557	0.6244	0.6964	0.6958
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.7141	0.7652	0.6628	0.7106	0.7497	0.6526	0.6482	0.6801	0.5903	0.6404	0.6781	0.5906
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.8766	0.3474	0.1933	0.8457	0.3490	0.2002	0.8296	0.3417	0.1954	0.7896	0.3213	0.1843
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.8757	0.9304	0.9275	0.8436	0.9003	0.8991	0.8544	0.9068	0.9031	0.8140	0.8656	0.8655
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.8592	0.8631	0.7723	0.7282	0.7454	0.6831	0.7346	0.7378	0.6648	0.7217	0.7292	0.6608
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8696	0.7883	0.6142	0.8291	0.7581	0.5942	0.7723	0.7027	0.5508	0.7463	0.6873	0.5413
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.8825	0.2253	0.1286	0.8418	0.2177	0.1234	0.8206	0.2131	0.1211	0.7720	0.2016	0.1143
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.8766	0.7247	0.6268	0.8494	0.7041	0.6099	0.8431	0.6967	0.5997	0.8140	0.6748	0.5838
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9239	0.6033	0.4448	0.8616	0.5584	0.4126	0.8308	0.5450	0.4037	0.8036	0.5254	0.3890
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9620	0.4915	0.3263	0.8759	0.4476	0.2977	0.8274	0.4318	0.2883	0.8025	0.4159	0.2780
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.9115	0.0647	0.0366	0.8855	0.0596	0.0336	0.8317	0.0572	0.0323	0.7628	0.0531	0.0300

8.6.3 Lead Time Forecasts

In chapter 6, the lead time forecasts for the INARMA models were presented. For INARMA(0,0) and INMA(1) processes, the lead time forecast is simply given by:

$$E\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = (l+1)\lambda$$

Equation 8-18

$$E\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = (l+1)(1+\beta)\lambda$$

Equation 8-19

For INAR(1) and INARMA(1,1) processes, the lead time forecasts are:

$$E\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha}Y_t + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right]$$

Equation 8-20

$$E\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha}Y_t + \frac{\lambda(1+\beta)}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right]$$

Equation 8-21

The results of sections 8.6.2.4 and 8.6.2.5 show that the accuracy of forecasts of an all-INAR(1) method are generally better than those of an all-INARMA(1,1) method even for INARMA(1,1) series. Therefore, in this section we use an all-INAR(1) method and compare the lead time forecasts of this method with those of benchmarks.

The results of comparing the MSE of INARMA with that of benchmark methods for INARMA(0,0), INMA(1), INAR(1) and INARMA(1,1) series are presented in Table 8-50 to Table 8-61. This includes both cases of $l = 3$ and $l = 6$. The results using MASE are presented in Appendix 8.I.

For INARMA(0,0) series, the results of Table 8-50 show that the all-INAR(1) lead time forecasts are better than the best benchmark in most of the cases (with an exception of $\lambda = 0.7, n = 24$). The same is true for INMA(1) series, with some exceptions for $n = 24$.

Table 8-50 $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ of lead-time forecasts ($l = 3$) for INARMA(0,0) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.7373	0.8307	0.8409	0.8726	0.9393	0.9457	0.9482	0.9880	0.9912	0.9448	0.9681	0.9695
$\lambda = 0.5$	0.9128	0.9879	0.9946	0.9573	0.9968	0.9992	0.9467	0.9804	0.9822	0.9097	0.9357	0.9367
$\lambda = 0.7$	0.9583	1.0130	1.0166	0.9541	0.9862	0.9872	0.9332	0.9606	0.9611	0.8914	0.9142	0.9143
$\lambda = 1$	0.9589	0.9878	0.9876	0.9361	0.9625	0.9620	0.9040	0.9265	0.9257	0.8601	0.8835	0.8830
$\lambda = 3$	0.9238	0.9077	0.8984	0.8642	0.8432	0.8338	0.8292	0.8141	0.8055	0.8021	0.7884	0.7803
$\lambda = 5$	0.9120	0.8519	0.8342	0.8657	0.8127	0.7967	0.8270	0.7755	0.7598	0.7897	0.7418	0.7272
$\lambda = 20$	0.9025	0.6417	0.5990	0.8521	0.6085	0.5675	0.8197	0.5886	0.5494	0.7793	0.5561	0.5189

Table 8-51 $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ of lead-time forecasts ($l = 3$) for INMA(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9226	1.0030	1.0104	0.9503	0.9928	0.9956	0.9429	0.9749	0.9765	0.9075	0.9372	0.9386
$\beta = 0.5, \lambda = 0.5$	0.9668	1.0400	1.0463	0.9292	0.9766	0.9800	0.9077	0.9468	0.9493	0.8554	0.8898	0.8918
$\beta = 0.9, \lambda = 0.5$	0.9745	1.0630	1.0708	0.9331	0.9877	0.9916	0.9030	0.9444	0.9469	0.8419	0.8803	0.8826
$\beta = 0.1, \lambda = 1$	0.9935	1.0339	1.0348	0.9341	0.9628	0.9627	0.9005	0.9262	0.9258	0.8557	0.8814	0.8812
$\beta = 0.5, \lambda = 1$	1.0054	1.0366	1.0366	0.9199	0.9444	0.9438	0.8925	0.9156	0.9148	0.8340	0.8579	0.8575
$\beta = 0.9, \lambda = 1$	1.0143	1.0469	1.0465	0.9382	0.9631	0.9621	0.9013	0.9218	0.9203	0.8360	0.8568	0.8557
$\beta = 0.1, \lambda = 3$	0.9407	0.9237	0.9139	0.8797	0.8710	0.8626	0.8487	0.8312	0.8223	0.8107	0.7981	0.7900
$\beta = 0.5, \lambda = 3$	0.9729	0.9451	0.9334	0.9029	0.8737	0.8627	0.8729	0.8508	0.8408	0.8190	0.7970	0.7873
$\beta = 0.9, \lambda = 3$	1.0083	0.9679	0.9541	0.9234	0.8801	0.8666	0.8841	0.8521	0.8401	0.8274	0.7961	0.7846
$\beta = 0.1, \lambda = 5$	0.9367	0.8801	0.8631	0.8710	0.8150	0.7988	0.8364	0.7885	0.7732	0.8075	0.7592	0.7443
$\beta = 0.5, \lambda = 5$	0.9695	0.9048	0.8853	0.8995	0.8289	0.8103	0.8683	0.8063	0.7887	0.8225	0.7612	0.7442
$\beta = 0.9, \lambda = 5$	0.9940	0.9124	0.8899	0.9071	0.8238	0.8029	0.8743	0.7985	0.7784	0.8297	0.7582	0.7392

Table 8-52 $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ of lead-time forecasts ($l = 3$) with smoothing parameter 0.2 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \lambda = 0.5$	0.9441	1.0142	1.0203	0.9422	0.9906	0.9942	0.9398	0.9719	0.9736	0.9012	0.9277	0.9289
$\alpha = 0.5, \lambda = 0.5$	0.9037	0.9525	0.9560	0.8704	0.9106	0.9131	0.8268	0.8611	0.8631	0.7714	0.8063	0.8085
$\alpha = 0.9, \lambda = 0.5$	0.8049	0.6984	0.6780	0.7157	0.6624	0.6473	0.6897	0.6391	0.6243	0.6219	0.5734	0.5605
$\alpha = 0.1, \lambda = 1$	0.9950	1.0303	1.0307	0.9124	0.9404	0.9403	0.9040	0.9314	0.9313	0.8533	0.8793	0.8792
$\alpha = 0.5, \lambda = 1$	0.9647	0.9843	0.9824	0.8941	0.9126	0.9110	0.8683	0.8820	0.8799	0.8117	0.8272	0.8254
$\alpha = 0.9, \lambda = 1$	0.7938	0.6568	0.6283	0.7219	0.6001	0.5749	0.6752	0.5758	0.5528	0.6244	0.5242	0.5024
$\alpha = 0.1, \lambda = 3$	0.9404	0.9214	0.9117	0.8876	0.8741	0.8649	0.8494	0.8370	0.8286	0.8052	0.7906	0.7825
$\alpha = 0.5, \lambda = 3$	0.9868	0.9416	0.9267	0.9043	0.8639	0.8504	0.8770	0.8385	0.8255	0.8203	0.7864	0.7745
$\alpha = 0.9, \lambda = 3$	0.7825	0.4666	0.4243	0.7136	0.4197	0.3827	0.6830	0.4084	0.3724	0.6364	0.3850	0.3508
$\alpha = 0.1, \lambda = 5$	0.9398	0.8872	0.8696	0.8577	0.8118	0.7964	0.8487	0.7989	0.7833	0.8011	0.7536	0.7389
$\alpha = 0.5, \lambda = 5$	0.9849	0.8914	0.8674	0.9053	0.8171	0.7949	0.8794	0.8054	0.7844	0.8246	0.7455	0.7253
$\alpha = 0.9, \lambda = 5$	0.7978	0.3639	0.3223	0.7284	0.3349	0.2968	0.6803	0.3201	0.2843	0.6294	0.3002	0.2666

Table 8-53 $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.5 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \lambda = 0.5$	0.8202	0.9889	1.0012	0.7103	0.8761	0.8946	0.7140	0.8540	0.8617	0.6728	0.8053	0.8140
$\alpha = 0.5, \lambda = 0.5$	0.8152	0.9551	0.9479	0.7505	0.8866	0.8825	0.7104	0.8328	0.8258	0.6566	0.7783	0.7749
$\alpha = 0.9, \lambda = 0.5$	1.0550	0.5086	0.3537	0.9908	0.5117	0.3575	0.9417	0.4827	0.3368	0.8686	0.4493	0.3156
$\alpha = 0.1, \lambda = 1$	0.7599	0.8681	0.8403	0.6779	0.7799	0.7567	0.6688	0.7718	0.7501	0.6283	0.7240	0.7033
$\alpha = 0.5, \lambda = 1$	0.8607	0.8891	0.8204	0.7999	0.8315	0.7695	0.7669	0.7907	0.7273	0.7220	0.7482	0.6894
$\alpha = 0.9, \lambda = 1$	1.0446	0.3595	0.2321	0.9967	0.3383	0.2190	0.9427	0.3302	0.2135	0.8687	0.2953	0.1905
$\alpha = 0.1, \lambda = 3$	0.6633	0.6147	0.5303	0.6263	0.5800	0.4968	0.6038	0.5614	0.4833	0.5732	0.5315	0.4577
$\alpha = 0.5, \lambda = 3$	0.8636	0.6624	0.5257	0.7882	0.6057	0.4822	0.7691	0.5909	0.4702	0.7130	0.5554	0.4434
$\alpha = 0.9, \lambda = 3$	1.0440	0.1492	0.0887	0.9923	0.1413	0.0843	0.9500	0.1361	0.0812	0.8741	0.1263	0.0751
$\alpha = 0.1, \lambda = 5$	0.6527	0.5102	0.4046	0.5950	0.4720	0.3774	0.5958	0.4671	0.3722	0.5596	0.4401	0.3514
$\alpha = 0.5, \lambda = 5$	0.8592	0.5293	0.3882	0.7813	0.4819	0.3539	0.7587	0.4735	0.3476	0.7183	0.4396	0.3219
$\alpha = 0.9, \lambda = 5$	1.0677	0.0967	0.0564	0.9947	0.0895	0.0522	0.9433	0.0866	0.0505	0.8708	0.0809	0.0472

The results of Table 8-53 for INAR(1) series reveal that, for high autoregressive parameters, the improvement of INARMA over the best benchmark is narrow for short length of history (in case of $n = 24$, INARMA is even worse). However with more observations, the improvement also increases. For small autoregressive parameters, the results of Table 8-52 show that INARMA always outperforms the benchmark methods. Again, the improvement increases with an increase in the length of history.

Table 8-54 $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.2 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9369	1.0169	1.0242	0.9479	0.9921	0.9951	0.9330	0.9650	0.9668	0.8867	0.9178	0.9195
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.0164	1.0937	1.1002	0.9385	0.9915	0.9953	0.8892	0.9311	0.9338	0.8316	0.8679	0.8700
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9728	1.0101	1.0118	0.8611	0.8924	0.8939	0.8120	0.8471	0.8490	0.7593	0.7894	0.7908
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.7710	0.6826	0.6646	0.7032	0.6297	0.6135	0.6573	0.6016	0.5875	0.6034	0.5641	0.5519
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9975	1.0335	1.0342	0.9151	0.9411	0.9409	0.9053	0.9308	0.9305	0.8528	0.8780	0.8779
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0291	1.0487	1.0470	0.9294	0.9539	0.9530	0.8891	0.9065	0.9048	0.8321	0.8514	0.8502
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9776	0.9953	0.9927	0.8819	0.8938	0.8909	0.8278	0.8396	0.8372	0.7817	0.7921	0.7897
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.7905	0.6207	0.5913	0.7148	0.5925	0.5671	0.6650	0.5543	0.5310	0.6202	0.5197	0.4979
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9627	0.9084	0.8908	0.8969	0.8410	0.8240	0.8528	0.7949	0.7788	0.8188	0.7691	0.7541
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9897	0.9024	0.8794	0.9067	0.8267	0.8062	0.8628	0.7857	0.7659	0.8240	0.7549	0.7362
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9816	0.8746	0.8481	0.8829	0.7884	0.7650	0.8442	0.7589	0.7369	0.7974	0.7113	0.6900
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.8126	0.3508	0.3100	0.7160	0.3352	0.2973	0.6811	0.3100	0.2747	0.6252	0.2905	0.2576

For INARMA(1,1) series, the results of Table 8-54 reveal that for small autoregressive parameters, INARMA outperforms the benchmark methods in most cases (except for sparse data and short history). The improvement increases with an increase in the length of history. For high autoregressive parameters, the improvement of INARMA over the best benchmark is narrow for short length of history (in case of $n = 24$, INARMA is even worse). However, with more observations, the improvement also increases.

Table 8-55 $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.5 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.7926	0.9844	1.0046	0.7387	0.8943	0.9066	0.7080	0.8508	0.8602	0.6609	0.8020	0.8151
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8284	1.0145	1.0283	0.7348	0.8943	0.9020	0.7039	0.8473	0.8523	0.6502	0.7844	0.7891
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.8944	1.0074	0.9783	0.7886	0.8929	0.8719	0.7309	0.8355	0.8170	0.6878	0.7789	0.7584
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.0603	0.5206	0.3644	0.9695	0.4808	0.3352	0.9265	0.4691	0.3282	0.8538	0.4430	0.3099
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.7832	0.8878	0.8586	0.6954	0.7909	0.7656	0.6774	0.7742	0.7505	0.6433	0.7359	0.7132
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.8336	0.8953	0.8471	0.7351	0.7975	0.7544	0.7160	0.7657	0.7200	0.6662	0.7168	0.6756
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9017	0.9086	0.8233	0.8102	0.8075	0.7288	0.7721	0.7688	0.6955	0.7151	0.7143	0.6467
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.0454	0.3365	0.2171	0.9817	0.3287	0.2121	0.9322	0.3129	0.2021	0.8667	0.2929	0.1888
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.6847	0.5329	0.4230	0.6441	0.4943	0.3906	0.6123	0.4713	0.3743	0.5900	0.4578	0.3637
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.7799	0.5191	0.3907	0.7282	0.4872	0.3678	0.6804	0.4559	0.3444	0.6462	0.4364	0.3297
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8976	0.5027	0.3589	0.8053	0.4581	0.3287	0.7748	0.4442	0.3189	0.7296	0.4119	0.2948
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.0675	0.0921	0.0538	1.0085	0.0898	0.0523	0.9440	0.0829	0.0483	0.8732	0.0781	0.0455

When $l = 6$, the results of Table 8-56 and Table 8-57 show that the improvement by using INARMA (an all-INAR(1) method) over the benchmarks is generally greater than the case of $l = 3$ for INARMA(0,0) and INMA(1) series. This is also true for INAR(1) and INARMA(1,1) series with small autoregressive parameters.

Table 8-56 $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ of lead-time forecasts $(l = 6)$ for INARMA(0,0) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.7796	0.9163	0.9315	0.7748	0.8717	0.8815	0.9112	0.9772	0.9824	0.9101	0.9488	0.9509
$\lambda = 0.5$	0.8894	1.0054	1.0161	0.8839	0.9565	0.9615	0.8798	0.9363	0.9394	0.8343	0.8782	0.8801
$\lambda = 0.7$	0.9194	1.0260	1.0341	0.9137	0.9761	0.9785	0.8802	0.9208	0.9212	0.7996	0.8416	0.8424
$\lambda = 1$	0.9390	1.0161	1.0193	0.8674	0.9197	0.9202	0.8225	0.8631	0.8623	0.7647	0.7959	0.7945
$\lambda = 3$	0.8688	0.8472	0.8330	0.7845	0.7530	0.7397	0.7503	0.7332	0.7213	0.6768	0.6616	0.6507
$\lambda = 5$	0.8601	0.7770	0.7533	0.7800	0.6999	0.6783	0.7253	0.6539	0.6336	0.6543	0.5956	0.5778
$\lambda = 20$	0.8767	0.5377	0.4903	0.7646	0.4681	0.4264	0.7239	0.4344	0.3956	0.6602	0.3997	0.3638

Table 8-57 $MSE_{INARMA} / MSE_{Benchmark}$ of lead-time forecasts $(l = 6)$ for INMA(1) series

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\beta=0.1, \lambda=0.5$	0.9555	1.0866	1.0992	0.9010	0.9718	0.9768	0.8826	0.9284	0.9306	0.8254	0.8694	0.8715
$\beta=0.5, \lambda=0.5$	0.9758	1.1192	1.1334	0.9130	0.9811	0.9859	0.8869	0.9499	0.9540	0.7785	0.8321	0.8355
$\beta=0.9, \lambda=0.5$	0.9807	1.1343	1.1492	0.9248	1.0144	1.0212	0.8610	0.9261	0.9304	0.7866	0.8389	0.8420
$\beta=0.1, \lambda=1$	0.9389	0.9944	0.9952	0.8628	0.9081	0.9082	0.8151	0.8529	0.8524	0.7649	0.7994	0.7987
$\beta=0.5, \lambda=1$	1.0045	1.0576	1.0578	0.8718	0.9157	0.9156	0.8293	0.8652	0.8644	0.7372	0.7696	0.7690
$\beta=0.9, \lambda=1$	1.0814	1.1128	1.1094	0.8983	0.9348	0.9332	0.8351	0.8787	0.8780	0.7429	0.7737	0.7725
$\beta=0.1, \lambda=3$	0.9124	0.8748	0.8592	0.8079	0.7739	0.7602	0.7394	0.7149	0.7030	0.6741	0.6553	0.6447
$\beta=0.5, \lambda=3$	0.9662	0.9205	0.9032	0.8329	0.8094	0.7957	0.7654	0.7379	0.7251	0.7011	0.6748	0.6630
$\beta=0.9, \lambda=3$	1.0009	0.9633	0.9456	0.8596	0.8297	0.8139	0.7917	0.7518	0.7364	0.7021	0.6679	0.6544
$\beta=0.1, \lambda=5$	0.9127	0.8509	0.8281	0.7848	0.7108	0.6897	0.7241	0.6617	0.6424	0.6698	0.6068	0.5887
$\beta=0.5, \lambda=5$	0.9715	0.8770	0.8489	0.8084	0.7152	0.6921	0.7664	0.6978	0.6771	0.6977	0.6281	0.6085
$\beta=0.9, \lambda=5$	0.9922	0.8905	0.8600	0.8249	0.7123	0.6872	0.7697	0.6726	0.6487	0.7043	0.6223	0.6010

Table 8-58 $MSE_{INARMA} / MSE_{Benchmark}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.2 for INAR(1) series

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\alpha=0.1, \lambda=0.5$	0.8625	0.9847	0.9967	0.9076	0.9769	0.9817	0.8948	0.9484	0.9515	0.8221	0.8695	0.8720
$\alpha=0.5, \lambda=0.5$	1.0555	1.1742	1.1849	0.9302	0.9968	1.0014	0.8466	0.9094	0.9138	0.7524	0.8044	0.8079
$\alpha=0.9, \lambda=0.5$	1.0075	0.9778	0.9611	0.8818	0.8539	0.8397	0.8450	0.8217	0.8080	0.7595	0.7278	0.7149
$\alpha=0.1, \lambda=1$	0.9882	1.0425	1.0431	0.8784	0.9175	0.9170	0.8314	0.8690	0.8684	0.7560	0.7902	0.7896
$\alpha=0.5, \lambda=1$	1.0573	1.0997	1.0987	0.9195	0.9563	0.9552	0.8406	0.8812	0.8809	0.7540	0.7849	0.7841
$\alpha=0.9, \lambda=1$	1.0224	0.8642	0.8311	0.8859	0.7750	0.7469	0.8428	0.7480	0.7224	0.7515	0.6479	0.6238
$\alpha=0.1, \lambda=3$	0.9045	0.8885	0.8752	0.8091	0.7854	0.7727	0.7458	0.7196	0.7073	0.6899	0.6730	0.6625
$\alpha=0.5, \lambda=3$	1.0213	0.9597	0.9405	0.8731	0.8446	0.8301	0.8273	0.8022	0.7887	0.7472	0.7167	0.7039
$\alpha=0.9, \lambda=3$	1.0113	0.6310	0.5780	0.8916	0.5759	0.5289	0.8439	0.5512	0.5061	0.7476	0.4875	0.4483
$\alpha=0.1, \lambda=5$	0.9003	0.8337	0.8110	0.7897	0.7115	0.6909	0.7370	0.6712	0.6516	0.6750	0.6152	0.5977
$\alpha=0.5, \lambda=5$	1.0195	0.9360	0.9091	0.8660	0.7777	0.7532	0.8137	0.7304	0.7077	0.7336	0.6596	0.6391
$\alpha=0.9, \lambda=5$	1.0231	0.5077	0.4525	0.8918	0.4638	0.4151	0.8473	0.4276	0.3828	0.7593	0.3885	0.3471

Table 8-59 $MSE_{INARMA} / MSE_{Benchmark}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.5 for INAR(1) series

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$
$\alpha=0.1, \lambda=0.5$	0.7025	0.9582	0.9857	0.6247	0.8199	0.8331	0.5745	0.7630	0.7799	0.5215	0.6905	0.7043
$\alpha=0.5, \lambda=0.5$	0.8524	1.1231	1.1434	0.7197	0.9131	0.9156	0.6479	0.8299	0.8336	0.5690	0.7232	0.7255
$\alpha=0.9, \lambda=0.5$	1.1450	0.7467	0.5466	1.0414	0.6744	0.4948	1.0005	0.6378	0.4654	0.8948	0.5661	0.4149
$\alpha=0.1, \lambda=1$	0.6911	0.8284	0.7860	0.5767	0.6950	0.6643	0.5338	0.6439	0.6137	0.4794	0.5823	0.5562
$\alpha=0.5, \lambda=1$	0.8340	0.9221	0.8532	0.7122	0.7868	0.7257	0.6494	0.7205	0.6662	0.5808	0.6405	0.5915
$\alpha=0.9, \lambda=1$	1.1542	0.4903	0.3295	1.0352	0.4408	0.2939	0.9963	0.4401	0.2945	0.8878	0.3718	0.2478
$\alpha=0.1, \lambda=3$	0.5518	0.5061	0.4185	0.4869	0.4383	0.3619	0.4381	0.3939	0.3249	0.4151	0.3808	0.3157
$\alpha=0.5, \lambda=3$	0.7784	0.5975	0.4667	0.6520	0.5253	0.4146	0.6223	0.4970	0.3921	0.5562	0.4413	0.3479
$\alpha=0.9, \lambda=3$	1.1314	0.2106	0.1279	1.0516	0.1967	0.1188	0.9973	0.1881	0.1133	0.8818	0.1691	0.1023
$\alpha=0.1, \lambda=5$	0.5334	0.4028	0.3086	0.4696	0.3467	0.2660	0.4351	0.3195	0.2432	0.4036	0.2995	0.2290
$\alpha=0.5, \lambda=5$	0.7723	0.4940	0.3607	0.6419	0.4011	0.2909	0.6079	0.3820	0.2778	0.5447	0.3420	0.2489
$\alpha=0.9, \lambda=5$	1.1593	0.1391	0.0817	1.0545	0.1297	0.0765	0.9927	0.1220	0.0720	0.8994	0.1079	0.0635

Table 8-60 $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.2 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9060	1.0221	1.0332	0.9114	0.9817	0.9867	0.8773	0.9416	0.9460	0.8151	0.8622	0.8649
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.1265	1.2635	1.2754	0.9392	1.0143	1.0196	0.8486	0.9209	0.9263	0.7638	0.8196	0.8232
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	1.1516	1.2325	1.2381	0.9878	1.0498	1.0535	0.8733	0.9306	0.9341	0.7564	0.8034	0.8062
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.0244	0.9352	0.9148	0.9012	0.8209	0.8031	0.8309	0.7946	0.7805	0.7339	0.7067	0.6946
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9968	1.0472	1.0475	0.8069	0.7319	0.7102	0.8184	0.8626	0.8631	0.7493	0.7859	0.7857
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0785	1.1185	1.1166	0.9066	0.9557	0.9558	0.8358	0.8781	0.8777	0.7402	0.7684	0.7670
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	1.1171	1.1825	1.1829	0.9376	0.9673	0.9650	0.8410	0.8700	0.8683	0.7467	0.7730	0.7715
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.0111	0.8583	0.8239	0.9020	0.7835	0.7542	0.8249	0.7473	0.7226	0.7269	0.6397	0.6174
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9241	0.8565	0.8332	0.8016	0.7424	0.7218	0.7275	0.6645	0.6456	0.6887	0.6289	0.6111
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9912	0.8566	0.8262	0.8560	0.7602	0.7341	0.7844	0.7027	0.6792	0.7024	0.6251	0.6044
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	1.0538	0.9423	0.9107	0.9001	0.7968	0.7694	0.8238	0.7342	0.7098	0.7444	0.6602	0.6380
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.0067	0.4899	0.4369	0.8862	0.4498	0.4017	0.8352	0.4233	0.3787	0.7291	0.3757	0.3360

Table 8-61 $\text{MSE}_{\text{INARMA}} / \text{MSE}_{\text{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.5 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.7042	0.9695	1.0084	0.6267	0.8313	0.8481	0.5555	0.7524	0.7762	0.5135	0.6838	0.7022
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8479	1.1332	1.1494	0.6814	0.8807	0.8870	0.5838	0.7769	0.7926	0.5222	0.6861	0.6938
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9535	1.1744	1.1651	0.8018	0.9757	0.9598	0.6933	0.8499	0.8383	0.6112	0.7413	0.7295
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.1333	0.7081	0.5236	1.0656	0.6303	0.4630	0.9797	0.6255	0.4586	0.8789	0.5565	0.4072
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.6951	0.8151	0.7736	0.4846	0.3506	0.2659	0.5338	0.6524	0.6276	0.4804	0.5845	0.5592
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.7753	0.8666	0.8034	0.6309	0.7326	0.6886	0.5860	0.6681	0.6238	0.5191	0.5816	0.5393
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8907	0.9756	0.8926	0.7496	0.7941	0.7204	0.6635	0.7086	0.6475	0.5940	0.6314	0.5747
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.1266	0.4673	0.3099	1.0550	0.4426	0.2935	0.9801	0.4372	0.2912	0.8728	0.3766	0.2513
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.5502	0.4146	0.3174	0.4870	0.3610	0.2743	0.4375	0.3241	0.2474	0.4218	0.3097	0.2358
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.6697	0.4177	0.3047	0.5673	0.3623	0.2632	0.5251	0.3368	0.2448	0.4660	0.3026	0.2219
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8193	0.4787	0.3404	0.7079	0.4066	0.2885	0.6455	0.3788	0.2698	0.5771	0.3386	0.2420
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.1254	0.1355	0.0798	1.0412	0.1238	0.0729	0.9907	0.1195	0.0705	0.8700	0.1045	0.0615

In general, the lead time forecasts produced by an all-INAR(1) method beat the benchmark methods except for the cases when the data is sparse and the sample is small. This can be attributed to the fact that there is less positive data available for estimation of parameters for an INAR(1) process. The improvement increases with an increase in the length of history. When the lead time increases, the improvement of INARMA over benchmarks generally increases except for the case where the autoregressive parameter is high.

8.7 Conclusions

In this chapter, the results of the simulation experiment have been presented. The performance of YW and CLS estimation methods (and CML for INAR(1) process) in terms of the accuracy of estimates and also their impact on forecast accuracy has been examined. The results show that when the length of history is short, CLS produces better forecasts than YW especially in the presence of a high autoregressive parameter. For cases where the autoregressive parameter is low, and also for the INMA(1) process, the two estimation methods are close. Also when the number of observations increases, the two estimation methods will produce close forecasts (in terms of MSE and MASE).

The Croston-SBA categorization (Syntetos et al., 2005) has been tested and validated for an i.i.d. Poisson process (an INARMA(0,0) process). Although the categorization was originally developed using MSE of forecasts, the results of simulation show that it also holds when MASE of forecasts are considered. The simulation results show that the Croston-SBA categorization also holds for INAR(1), INMA(1) and INARMA(1,1) processes.

It has been found that when the number of observation increases, the advantage of SBA over Croston in terms of MSE decreases until it reaches a limit.

Four INARMA processes have been used in this study: INARMA(0,0), INAR(1), INMA(1) and INARMA(1,1). The identification is therefore limited to selecting the best process among them. As discussed in chapter 4, two identification procedures are used. A two-stage identification procedure first uses the Ljung-Box statistic to distinguish between INARMA(0,0) and other processes. The AIC is then used to select among the other INARMA models. A one-stage identification procedure only uses AIC to select among all INARMA models including INARMA(0,0). The results show that the two-stage method provides better results for the INARMA(0,0) model (in terms of the percentage of series for which the correct model is identified). However, for other models, the one-stage method produces better results. In terms of the accuracy of forecasts, for an INARMA(0,0) process, the two-stage method produces better forecasts using MSE and MASE in most of the cases. For other processes, when the autoregressive parameter is high, the one-stage method produces much better forecasts. This is also true for high moving average parameters but the difference is smaller. When more observations are available, the two methods produce similar forecasts. The results also suggest that, as expected, misidentification has a high effect on forecast accuracy when the autoregressive parameter is high. But for MA processes and for AR processes with low autoregressive parameters, the effect of misidentification is not high.

As a potential substitute to identification, the most general INARMA model can be used. For example, if data is in fact an INAR(1) process, the estimated MA parameter should be close to zero. The results show that, in the presence of an AR component, using the most general INARMA model results in more accurate forecasts (in terms of MSE and MASE) than those of identification, especially for short history and high AR parameter. When the number of observations increases, the results of two methods will be close.

We have also tested using an INAR(1) method to forecast all four INARMA models and have compared the results to the case of using an INARMA(1,1) method for all four models. The results show that the all-INAR(1) method generally produces better forecasts than the all-INARMA(1,1) method even for MA series.

The INARMA forecasts are compared to those of Croston, SBA and SBJ. For INARMA(0,0) and INMA(1) processes, the improvement by using INARMA over benchmark methods is small. But when data is produced by INAR(1) or INARMA(1,1) and the autoregressive parameter is high, INARMA produces much more accurate one-step ahead forecasts than the benchmark methods. The degree of improvement generally increases when more observations are available.

The results for three-step and six-step ahead forecasts show that, for INMA(1), and INAR(1) and INARMA(1,1) processes with small autoregressive parameters, the forecast accuracy of INARMA over benchmarks is improved compared to the one-step ahead forecasts. However, the same is not true for INARMA processes with high autoregressive parameters.

Finally, the lead time forecasts of an all-INAR(1) method have been compared to those of the benchmark methods. The results show that the all-INAR(1) method generally beats the benchmark methods and the improvement increases when more observations are available. The only exception is when both autoregressive parameter and lead time are high and the number of observations is small. In that case, the best benchmark method outperforms the all-INAR(1) method. Even for such cases, when the number of observations increases, INARMA starts to produce more accurate lead time forecasts than benchmarks.

Chapter 9 EMPIRICAL ANALYSIS

9.1 Introduction

As discussed in chapter 3, INARMA models have been developed for forecasting count data. The application areas of these models have been mainly for counts of events or individuals such as the number of patients in a hospital's emergency unit each hour.

Intermittent series, as *a series of non-negative integer values where some values are zero* (Shenstone and Hyndman, 2005), can be considered as a special class of count series. However, there is no empirical evidence on the performance of INARMA models in this area.

Accurate demand forecasting is a key to better inventory management. Although this research only focuses on forecasting, an improvement in accuracy of forecasts can be translated to fewer inventories, less obsolescence and better customer service.

This PhD thesis has suggested using INARMA models to forecast intermittent demand. The performance of these models has been compared to that of the benchmark methods of Croston, SBA and SBJ in Chapter 8. However, it was assumed that data were produced by one of the four INARMA models: INARMA(0,0), INAR(1), INMA(1) or INARMA(1,1). In this chapter, the model assumptions are relaxed by testing the results on empirical data.

This chapter is organized as follows. The purposes of empirical analysis are explained in section 9.2. The series for this study consist of the demand data of 16,000 Royal Air Force (RAF) SKUs, some of which are highly lumpy, and 3,000 data series from the automotive industry. The filtering mechanism applied to the demand data for INARMA forecasting is discussed in section 9.3. Details of empirical analysis design are provided in section 9.4. The accuracy of INARMA forecasts and those of the benchmark methods are compared in section 9.5. In this section, the results of identification are compared with treating all as INAR(1) and INARMA(1,1). The h-step ahead forecasts and lead time forecasts are also presented. The sensitivity of the results to the length of history has been tested. Finally, the conclusions of the empirical analysis are given in section 9.6.

9.2 Rationale for Empirical Analysis

The main purpose of empirical analysis is to validate the theoretical and simulation findings on real data. The results of simulation show that when data is produced by INAR(1) or INARMA(1,1) and the autoregressive parameter is high, INARMA forecasting methods produce much more accurate forecasts than benchmark methods. But when data is produced by INARMA(0,0) or INMA(1), the improvement by using INARMA over benchmark methods is small. We are interested in finding out whether the INARMA forecasting approach still outperforms the benchmark methods for real intermittent demand data.

In the simulation chapter we also looked at the effect of various factors on forecast accuracy. The effect of YW and CLS (and CML for INAR(1)) estimates on the accuracy of INARMA forecasts was studied. The results of identification or using the most general INARMA model in the class were compared. The sample size effect on the accuracy of forecasts was also tested. The empirical analysis will assess the effect of these factors and enable us to validate the simulation results.

9.3 Demand Data Series

The real demand data series for this research consists of the Royal Air Force (RAF) individual demand histories of 16,000 SKUs over a period of 6 years (monthly observations). We have also used another data set which consists of 3,000 real intermittent demand data series from the automotive industry[1] (from Syntetos and Boylan, 2005) which, unlike the previous one, has more occurrences of positive demand than zeros. This data series consists of demand histories of 3,000 SKUs over a period of 2 years (24 months). These two data sets are called 16,000 and 3,000 series from now on.

The 16,000 series are useful in assessing the effect of length of history on the accuracy of forecasts because it has longer history. However, it is can be categorized as a set of slower intermittent series because it has many periods of no demand. On the other hand, the 3,000 series has a very short history but it contains faster intermittent series with more positive demands.

As previously mentioned, this research has focused on INARMA processes with Poisson innovations. Although some of the theoretical results are not based on a distributional assumption, whenever a specific distribution was needed, such as for estimation of parameters, a Poisson distribution was assumed.

Out of the four processes of this study, three of them have a Poisson distribution when the marginal distribution is Poisson. The only exception is the INARMA(1,1) process where:

$$Y_t = \alpha \circ Y_{t-1} + Z_t + \beta \circ Z_{t-1}$$

Equation 9-1

[1] This data set is available from: http://www.forecasters.org/ijf/data/Empirical%20Data.xls

$$E(Y_t) = \frac{1+\beta}{1-\alpha}\lambda$$

Equation 9-2

$$\text{var}(Y_t) = \frac{1}{1-\alpha^2}[1+\alpha+\beta+3\alpha\beta]\lambda$$

Equation 9-3

$$\frac{\text{var}(Y_t)}{E(Y_t)} = \frac{1+\alpha+\beta+3\alpha\beta}{1+\alpha+\beta+\alpha\beta} \leq 1.5 \qquad \text{for } 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1$$

Equation 9-4

In order to remove the data series with highly variable demands, a Poisson dispersion test (also called the variance test) is needed for all processes except INARMA(1,1). Under the null hypothesis that $X_1, \dots, X_n$ are Poisson distributioned, the test statistic:

$$T_{CC} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\bar{X}}$$

Equation 9-5

has a chi-square distribution with $(n-1)$ degrees of freedom. Therefore, H_0 is rejected if $T_{CC} > \chi^2_{n-1;1-\alpha}$.

A revised statistic is used to allow for the difference between the mean and variance of an INARMA(1,1) process. The new test statistic is given by:

$$T_{CCR} = \frac{T_{CC}}{1.5}$$

Equation 9-6

The new statistic also has a chi-square distribution with $(n-1)$ degrees of freedom.

Further filtering of data was performed for series with fewer than two nonzero demands. As previously mentioned in Chapter 8, the benchmark methods need at least two nonzero observations for initialization.

Out of the 16,000 series, 5,168 series met the above criteria and therefore are used for empirical analysis. The filtering of the 3,000 series results in 1,943 series. It can be seen that although a substantial number of series has the potential to benefit from PoINARMA models, for a large number of series these models are not appropriate.

Other distributional assumptions would obviously result in different number of filtered series, which can be pursued as a further study.

9.4 Design of Empirical Analysis

The design of the empirical analysis follows the detailed simulation design of Chapter 7. As discussed in section 7.3.3.4, two fixed values have been used for the smoothing parameter of Croston, SBA, and SBJ methods ($\alpha = 0.2, 0.5$). The initialization for these methods is based on using the first inter-demand interval as the first smoothed inter-demand interval and the average of the first two non-zero observations as the first smoothed size.

The data series is divided into two parts: "estimation period" for initialization and estimation of parameters and "performance period" for assessing the accuracy of forecasts. If at least two non-zero demands are observed in the estimation period, the first half of the observations is assigned for the estimation period and the other half for the performance period. However, if fewer than two non-zero demands are observed in the estimation period, this period will be extended until the second non-zero demand is observed. When the effect of length of history is tested, the performance period is fixed and the estimation period varies.

It is assumed that there are four possible INARMA models to use for forecasting. Therefore, identification is undertaken among these models. This is done by applying both two-stage and one-stage identification procedures (see section 4.6). The former first uses the Ljung-Box test to distinguish between INARMA(0,0) and the other three models. Then, the AIC is used to select among the other models. The latter uses only the AIC to select among all INARMA models.

Based on the simulation results of chapter 8, we suggested that general models (INAR(1) and INARMA(1,1) were tested in chapter 8) can be used as alternatives to identification. This is also tested on empirical data.

YW and CLS (and CML only for the INAR(1) model) have been used to estimate the parameters of INARMA models. The simulation results show that these methods result in similar forecasts when the length of history is high. But for short history and high autoregressive parameters, CLS generally produces more accurate one-step ahead forecasts in terms of MSE and MASE. These estimation methods have been used for empirical analysis to test the validity of the simulation findings. It is worth mentioning that the INARMA forecasts are minimum mean square (MMSE). Therefore, we expect an improvement in terms of MSE but not necessarily in terms of MASE.

The accuracy measures used in empirical analysis are discussed in chapter 2 in detail. In addition to the measures used in the simulation experiment (ME, MSE, and MASE), two relative-to-another-method measures have been used. As previously mentioned in section 2.4, the percent better (PB) determines how often a method is better than another method. However, it does not show how much the improvement is. The relative geometric root-mean-square error (RGRMSE) is used to calculate the magnitude of improvement of one method over another.

9.5 INARMA vs Benchmark Methods

In this section, the results of comparing INARMA forecasts with those of benchmarks are presented. First, we use an all-INAR(1) approach assuming that all the series are in fact INAR(1) processes. The different estimation methods used in the simulation chapter for estimation of INAR(1) parameters are also used here. An all-INARMA(1,1) approach is then used and the results are compared to both benchmark methods and all-INAR(1) results. As previously mentioned in sections 8.6.2.2 and 8.6.2.4, these approaches especially perform well compared to identification for highly auto-correlated data and short data histories. To assess their performance for empirical data, the forecast accuracy results based on identification are also presented.

The results of identification show what INARMA models each series follows. Those series that follow a specific INARMA model are then separated and each one is forecasted with the corresponding INARMA method (either INARMA(0,0), INAR(1), INMA(1), or INARMA(1,1)). The accuracy of these INARMA forecasts is compared to the corresponding simulation results. The results include one-step, three-step and six-step-ahead forecasts.

The sensitivity of INARMA forecasts to the length of history is also tested for the 16,000 series. Finally, the results of lead time forecasts for the INAR(1) model are presented for both the 16,000 and 3,000 series.

9.5.1 All-INAR(1)

We suggest in chapter 8 that using a general INARMA model produces forecasts comparable to those based on identification. As we will show in section 9.5.3, most of the series in both 16,000 and 3,000 series are identified as INARMA(0,0) or INAR(1). Therefore, using INAR(1) to forecast seems to be a promising approach for these datasets. The results of INAR(1) and benchmarks for all points in time and issue points are shown in Table 9-1 and Table 9-2 for 16,000 series.

Recall from chapter 5 that, as pointed out by Al-Osh and Alzaid (1987), for small sample sizes ($n \leq 75$) and small autoregressive parameter ($\alpha = 0.1$), because the sample contains many zero values, CML is not as good as YW in terms of bias and MSE of the estimates. The simulation results in section 8.3 show that for corresponding parameters (see section 9.5.4), YW yields better forecasts than CLS and CML using MSE and MASE. We test this by presenting CLS, YW and CML based forecasts. The empirical results confirm the corresponding simulation results.

Table 9-1 Comparing INAR(1) with benchmarks for all points in time (16000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.0719	-0.0933	-0.0402	-0.0086	-0.0367	0.0196	0.0261	0.0283	0.0282
MSE	0.3910	0.4205	0.3802	0.3859	0.3793	0.3846	0.3609	0.3527	0.3555
MASE	2.8594	2.8852	2.7051	2.4925	2.6881	2.3640	1.9789	1.9124	1.9432
PB of MASE INAR(1)-CLS/Benchmark	0.6830	0.6952	0.6302	0.5624	0.6236	0.5116			
PB of MASE INAR(1)-YW/Benchmark	0.6742	0.6868	0.6216	0.5555	0.6150	0.5050			
PB of MASE INAR(1)-CML/Benchmark	0.6651	0.6878	0.6116	0.5545	0.6050	0.5050			
RGRMSE INAR(1)-CLS/Benchmark	0.7812	0.7401	0.8272	0.9024	0.8352	0.9927			
RGRMSE INAR(1)-YW/Benchmark	0.7898	0.7493	0.8382	0.9157	0.8463	1.0077			
RGRMSE INAR(1)-CML/Benchmark	0.7894	0.7494	0.8367	0.9152	0.8449	1.0075			

Table 9-2 Comparing INAR(1) with benchmarks for issue points (16000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.0266	-0.0441	0.0028	0.0339	0.0061	0.0598	-0.0598	-0.0324	-0.0197
MSE	0.5329	0.5666	0.5237	0.5344	0.5231	0.5359	0.5792	0.5218	0.5420
MASE	0.3309	0.3375	0.3188	0.3048	0.3175	0.2951	0.3679	0.3473	0.3444
PB of MASE INAR(1)-CLS/Benchmark	0.4590	0.4509	0.4016	0.3219	0.3949	0.2900			
PB of MASE INAR(1)-YW/Benchmark	0.4584	0.4493	0.4026	0.3273	0.3956	0.2941			
PB of MASE INAR(1)-CML/Benchmark	0.4826	0.4785	0.4234	0.3534	0.4166	0.3196			
RGRMSE INAR(1)-CLS/Benchmark	1.2416	1.3295	1.3402	1.6694	1.3531	1.8393			
RGRMSE INAR(1)-YW/Benchmark	1.1734	1.2477	1.2728	1.5709	1.2847	1.7301			
RGRMSE INAR(1)-CML/Benchmark	1.1060	1.2158	1.1989	1.5367	1.2105	1.6960			

The results of Table 9-1 show that INAR(1) produces better results than benchmarks regardless of the estimation method. The results also show that YW based INAR(1) is the best estimation method among the three methods, outperforming the best benchmark method (SBJ 0.2 for MSE and SBJ 0.5 for MASE) by 7 percent in terms of MSE and 19 percent in terms of MASE. The results of PB and RGRMSE also show superior performance of INARMA compared to the benchmark methods.

As can be seen from Table 9-2, when only issue points are considered, the INARMA forecasts are biased, agreeing with the simulation results. We find that this is true for all the cases that only issue points are considered so we do not discuss this again.

It can be seen from Table 9-1 that the MASE of all methods is very high which, as explained in section 2.4, suggests that all of these methods are worse than naïve. The reason is that because the data series contain many zeros in the estimation period, the error of naïve in most of the periods is zero. Therefore the in-sample MAE is very small and the MASE is very large.

However, when only issue points are considered, because it is likely that a nonzero demand is followed with a zero demand, the absolute error of naïve and therefore the in-sample MAE is large. As a result, the MASE of the forecasting methods is smaller compared to the all points in time case. The results of Table 9-2 confirm this.

As a result, for highly intermittent data, MASE does not provide reliable results for all points in time and issue points.

The simulation results of section 8.6.1 show the superiority of INARMA(0,0) forecasts compared to the benchmarks for $\lambda = 0.3, n = 96$. It will be seen in section 9.5.4 that the majority of 16,000 series are identified as INARMA(0,0) with λ close to 0.3 (see appendix 9.A) so the results for 16,000 series for all points in time agree with the simulation results.

The results for 3,000 series are shown in Table 9-3 and Table 9-4. It can be seen that YW again results in more accurate forecasts than both CLS and CML. CLS and CML results are worse than benchmarks and YW results only improve MSE by 0.5 percent compared to the best benchmark which is SBA 0.2. However, the MASE of SBJ 0.2 is better than that of INAR(1)-YW by one percent. As mentioned in chapter 6, the INARMA forecasts provide the minimum MSE and not MASE.

The results of Table 9-3 also show the superior performance of INAR(1)-YW to the benchmark methods in terms of PB and RGRMSE except for SBA and SBJ 0.2.

Based on the results of Table 9-4, INAR(1)-YW is better than the best benchmark in terms of MSE although the INARMA forecasts are biased when only issue points are considered. However, the MASE is still slightly worse.

Table 9-3 Comparing INAR(1) with benchmarks for all points in time (3000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.0419	-0.1016	0.1662	0.4336	0.1894	0.6120	0.1016	0.0136	0.0145
MSE	3.2574	3.7054	3.2483	3.6470	3.2550	3.8249	3.3540	3.2319	3.2640
MASE	0.8694	0.9277	0.8543	0.8848	0.8535	0.8914	0.8757	0.8636	0.8720
PB of MASE INAR(1)-CLS/Benchmark	0.5097	0.5528	0.4885	0.5211	0.4886	0.5329			
PB of MASE INAR(1)-YW/Benchmark	0.5171	0.5551	0.4842	0.5211	0.4839	0.5295			
PB of MASE INAR(1)-CML/Benchmark	0.5133	0.5522	0.4828	0.5136	0.4812	0.5259			
RGRMSE INAR(1)-CLS/Benchmark	0.5329	0.9814	0.9370	1.0027	1.0037	1.0067			
RGRMSE INAR(1)-YW/Benchmark	0.9471	0.9175	0.9765	0.9868	0.9826	0.9857			
RGRMSE INAR(1)-CML/Benchmark	0.9476	0.9207	0.9791	0.9898	0.9847	0.9894			

Table 9-4 Comparing INAR(1) with benchmarks for issue points (3000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.0518	-0.1058	0.1544	0.4233	0.1773	0.5996	-0.0037	-0.0650	-0.0732
MSE	3.3807	3.8368	3.3685	3.7724	3.3749	3.9483	3.4687	3.3563	3.3865
MASE	0.9083	0.9686	0.8918	0.9210	0.8907	0.9260	0.9164	0.9054	0.9130
PB of MASE INAR(1)-CLS/Benchmark	0.5087	0.5555	0.4861	0.5157	0.4857	0.5231			
PB of MASE INAR(1)-YW/Benchmark	0.5117	0.5539	0.4756	0.5127	0.4744	0.5161			
PB of MASE INAR(1)-CML/Benchmark	0.5097	0.5519	0.4741	0.5076	0.4723	0.5162			
RGRMSE INAR(1)-CLS/Benchmark	1.0187	0.9738	1.0354	1.0442	1.0434	1.0503			
RGRMSE INAR(1)-YW/Benchmark	0.9927	0.9618	1.0200	1.0311	1.0293	1.0377			
RGRMSE INAR(1)-CML/Benchmark	0.9953	0.9686	1.0291	1.0434	1.0394	1.0495			

The simulation results of chapter 8 suggest that, with the presence of a high autocorrelation, INAR(1) outperforms the benchmarks. Also, the simulation results show that for $\alpha = 0.1$ and $\lambda = 1, 3$, when the number of observations is small ($n = 24$), SBA and SBJ (with smoothing parameter 0.2) are better than INAR(1) in terms of both MSE and MASE. But for higher number of observations, INAR(1) starts to slightly perform better.

It will be seen in section 9.5.4 that the estimated autoregressive parameter is generally close to 0.1. Therefore, the results for 3000 series agree with the simulation results.

9.5.2 All-INARMA(1,1)

In this section it is assumed that all series follow the INARMA(1,1) model. The results for 16,000 series are presented in Table 9-5 and Table 9-6.

The simulation results show that for INARMA(1,1) processes, CLS produces better results than YW. We test this for empirical data by presenting both CLS and YW-based forecasts. As explained in chapter 5, the CML estimates have only been obtained for INAR(p) models and therefore these estimates are not presented for the INARMA(1,1) model.

The results of Table 9-5 show that INARMA improves the accuracy of forecast by 3 percent in terms of MSE and 19 percent in terms of MASE compared to the best benchmark (SBJ 0.2 for MSE and SBJ 0.5 for MASE). The RGRMSE results confirm the MSE results. It can be seen that the YW-based INARMA(1,1) forecasts are worse than CLS-based forecasts, agreeing with simulation results.

Table 9-5 Comparing INARMA(1,1) with benchmarks for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.0719	-0.0933	-0.0402	-0.0086	-0.0367	0.0196	0.0258	-0.0039
MSE	0.3910	0.4205	0.3802	0.3859	0.3793	0.3846	0.3668	0.4833
MASE	2.8594	2.8852	2.7051	2.4925	2.6881	2.3640	1.9128	2.3085
PB of MASE INARMA-CLS/Benchmark	0.6692	0.6842	0.6161	0.5508	0.6095	0.5004		
PB of MASE INARMA-YW/Benchmark	0.7359	0.7413	0.7150	0.6922	0.7122	0.6691		
RGRMSE INARMA-CLS/Benchmark	0.7653	0.7294	0.8128	0.8918	0.8205	0.9818		
RGRMSE INARMA-YW/Benchmark	0.7843	0.7464	0.8200	0.8960	0.8285	0.9881		

Table 9-6 Comparing INARMA(1,1) with benchmarks for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.0266	-0.0441	0.0028	0.0339	0.0061	0.0598	-0.0326	-0.5620
MSE	0.5329	0.5666	0.5237	0.5344	0.5231	0.5359	0.5766	1.5057
MASE	0.3309	0.3375	0.3188	0.3048	0.3175	0.2951	0.3509	0.7313
PB of MASE INARMA-CLS/Benchmark	0.4958	0.4797	0.4349	0.3507	0.4275	0.3162		
PB of MASE INARMA-YW/Benchmark	0.3004	0.2923	0.2779	0.2375	0.2750	0.2185		
RGRMSE INARMA-CLS/Benchmark	1.0833	1.1712	1.1751	1.4768	1.1862	1.6282		
RGRMSE INARMA-YW/Benchmark	2.1539	2.3492	2.4021	2.9709	2.4289	3.3006		

The results of comparing INARMA(1,1) and benchmarks for all points in time and issue points for 3,000 series are presented in Table 9-7 and Table 9-8 for both CLS and YW estimates. The results confirm that INARMA(1,1) forecasts based on CLS estimates are better than those based on YW estimates.

The results of Table 9-7 show that both CLS and YW based INARMA(1,1) forecasts are worse than benchmarks. In order to understand the reason, we need to identify the autoregressive and moving average order of the data series. This will be examined in the next section.

Table 9-7 Comparing INARMA(1,1) with benchmarks for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.0419	-0.1016	0.1662	0.4336	0.1894	0.6120	-0.0461	0.0091
MSE	3.2574	3.7054	3.2483	3.6470	3.2550	3.8249	3.4038	4.0726
MASE	0.8694	0.9277	0.8543	0.8848	0.8535	0.8914	0.8869	0.9548
PB of MASE INARMA-CLS/Benchmark	0.5063	0.5401	0.4777	0.5125	0.4767	0.5209		
PB of MASE INARMA-YW/Benchmark	0.4734	0.5018	0.4517	0.4782	0.4522	0.4899		
RGRMSE INARMA-CLS/Benchmark	0.9893	0.9536	1.0231	1.0293	1.0289	1.0276		
RGRMSE INARMA-YW/Benchmark	0.7435	0.7078	0.7675	0.7680	0.7723	0.7669		

Table 9-8 Comparing INARMA(1,1) with benchmarks for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.0518	-0.1058	0.1544	0.4233	0.1773	0.5996	-0.1366	-0.1972
MSE	3.3807	3.8368	3.3685	3.7724	3.3749	3.9483	3.5569	4.1304
MASE	0.9083	0.9686	0.8918	0.9210	0.8907	0.9260	0.9327	0.9902
PB of MASE INARMA-CLS/Benchmark	0.5006	0.5378	0.4691	0.5027	0.4677	0.5097		
PB of MASE INARMA-YW/Benchmark	0.4805	0.5108	0.4558	0.4844	0.4570	0.4961		
RGRMSE INARMA-CLS/Benchmark	1.0343	0.9940	1.0629	1.0707	1.0712	1.0772		
RGRMSE INARMA-YW/Benchmark	0.8166	0.7749	0.8417	0.8469	0.8486	0.8527		

The simulation results for the corresponding parameter set ($\alpha = 0.1, 0.3$, β close to zero, and $\lambda = 1,3$) show that when the number of observations is small ($n = 24$), INARMA(1,1) performance is poor. But for a higher autoregressive parameter and number of observations, INARMA(1,1) performance improves greatly.

Comparing the results of this section with those of the previous section shows that treating all as INAR(1) produces better results than treating all as INARMA(1,1) which confirms the simulation results (see section 8.6.2.4). A possible explanation is that the number of parameters to be estimated and therefore the estimation error is less for all-INAR(1) compared to all-INARMA(1,1). The results of identification in the next section will show that 98.78 percent of 16,000 series (78.43 for 3,000 series) are identified as INAR(0,0) or INAR(1). This could also justify the superior performance of all-INAR(1) because only 1.22 percent of the series (21.57 for 3000 series) are either INMA(1) or INARMA(1,1). Although more series are identified as INMA(1) and INARMA(1,1) in 3,000 series, their moving average parameter is generally between zero and 0.1 and these processes are close to INARMA(0,0) and INAR(1). This could also help to understand why all-INAR(1) works better than all-INARMA(1,1) for 3,000 series.

9.5.3 Identification among four Processes

In this section, the appropriate INARMA model is identified among the four possible candidates. Both one-stage and two-stage identification procedures are tested (see section 4.6 for details). For both identification methods, the results for the case where all the INARMA forecasts are based on CLS and YW estimates are presented.

The accuracy of INARMA forecasts based on identification and treating all as INAR(1) or INARMA(1,1) for 16,000 series are compared in Table 9-9 and Table 9-10. As expected, identification (two-stage) produces slightly better results in terms of MSE and MASE, (except for MSE of all-INAR(1) for all points in time) but the results are generally close.

Table 9-9 The effect of identification on INARMA forecasts for all points in time (16000 series)

The identification method	ME	MSE	MASE	PB of MASE (INARMA/Benchmark)						RGRMSE (INARMA/Benchmark)					
				Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5
Two-stage identification (CLS)	0.0334	0.3529	1.8576	0.6659	0.6818	0.6099	0.5440	0.6030	0.4927	0.8004	0.7614	0.8514	0.9328	0.8595	1.0264
Two-stage identification (YW)	0.0321	0.3521	1.8647	0.6655	0.6811	0.6096	0.5440	0.6027	0.4924	0.7968	0.7577	0.8474	0.9286	0.8555	1.0217
One-stage identification (CLS)	0.0307	0.3581	1.9008	0.6665	0.6832	0.6123	0.5481	0.6054	0.4963	0.7948	0.7549	0.8443	0.9232	0.8524	1.0165
One-stage identification (YW)	0.0309	0.3533	1.9085	0.6359	0.6413	0.6150	0.5422	0.6022	0.4961	0.7945	0.7557	0.8436	0.9221	0.8529	1.0151
All-INAR(1) (YW)	0.0283	0.3527	1.9124	0.6742	0.6868	0.6216	0.5555	0.6150	0.5050	0.7898	0.7493	0.8382	0.9157	0.8463	1.0077
All-INARMA(1,1) (CLS)	0.0258	0.3668	1.9128	0.6692	0.6842	0.6161	0.5508	0.6095	0.5004	0.7653	0.7294	0.8128	0.8918	0.8205	0.9818

Table 9-10 The effect of identification on INARMA forecasts for issue points (16000 series)

The identification method	ME	MSE	MASE	PB of MASE (INARMA/Benchmark)						RGRMSE (INARMA/Benchmark)					
				Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5
Two-stage identification (CLS)	0.0351	0.5045	0.3055	0.5562	0.5246	0.4903	0.3836	0.4817	0.3448	0.9315	0.9659	1.0093	1.2102	1.0186	1.3311
Two-stage identification (YW)	0.0256	0.5187	0.3093	0.5514	0.5196	0.4856	0.3795	0.4768	0.3406	0.9587	0.9910	1.0392	1.2435	1.0488	1.3688
One-stage identification (CLS)	0.0012	0.5288	0.3293	0.5211	0.5014	0.4581	0.3682	0.4497	0.3308	1.0188	1.0941	1.1053	1.3783	1.1158	1.5188
One-stage identification (YW)	0.0220	0.5257	0.3313	0.5204	0.5023	04579	0.3575	0.4450	0.3185	1.0145	1.0722	1.0951	1.3723	1.1116	1.5252
All-INAR(1) (YW)	-0.0324	0.5218	0.3473	0.4584	0.4493	0.4026	0.3273	0.3956	0.2941	1.1734	1.2477	1.2728	1.5709	1.2847	1.7301
All-INARMA(1,1) (CLS)	-0.0326	0.5766	0.3509	0.4958	0.4797	0.4349	0.3507	0.4275	0.3162	1.0833	1.1712	1.1751	1.4768	1.1862	1.6282

The results of Table 9-9 and Table 9-10 also show that the two-stage identification procedure provides better results than the one-stage. Since the majority of series are identified as INARMA(0,0), this agrees with the simulation results of section 8.6.2.3.

It can also be seen that the CLS and YW yield close results for both two-stage and one-stage identification methods. In order to be consistent with the simulation analysis (see conclusion of section 8.4), we focus on the CLS-based two-stage identification results.

For 16,000 time series, out of the 5,168 series, 98.12 percent were identified as INARMA(0,0), 0.66 percent as INAR(1), 1.04 percent as INMA(1), and 0.17 percent were identified as INARMA(1,1).

As can be seen, the majority of the series are identified as INARMA(0,0). The simulation results show that when data is in fact INARMA(0,0), the all-INARMA(1,1) forecasts are close to INARMA(0,0) forecasts. The simulation results also show that when the order is known to be (0,0), all-INARMA(1,1) produces better forecasts than the best benchmark method. The results of Table 9-5 and Table 9-6 show that all-INARMA(1,1) for 16,000 series performs better than benchmarks, agreeing with the simulation results.

Now, the accuracy of INARMA forecasts based on identification and treating all as INAR(1) or INARMA(1,1) for 3,000 series are compared in Table 9-11 and Table 9-12. Here again, the two-stage identification produces slightly better results not only in terms of MSE and MASE, but also in terms of PB of MASE and RGRMSE (with the exception of MSE of all-INAR(1)). It should be mentioned that these results are still not as good as SBA and SBJ with smoothing parameter 0.2.

Table 9-11 The effect of identification on INARMA forecasts for all points in time (3000 series)

The identification method	ME	MSE	MASE	PB of MASE (INARMA/Benchmark)						RGRMSE (INARMA/Benchmark)					
				Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5
Two-stage identification (CLS)	0.0813	3.2925	0.8608	0.5238	0.5557	0.4973	0.5286	0.4974	0.5358	0.9376	0.9050	0.9599	0.9710	0.9647	0.9707
Two-stage identification (YW)	0.0191	3.2840	0.8603	0.5229	0.5514	0.4895	0.5244	0.4893	0.5315	0.9135	0.8875	0.9412	0.9544	0.9475	0.9500
One-stage identification (CLS)	0.0276	3.3088	0.8697	0.5156	0.5497	0.4852	0.5191	0.4854	0.5276	0.9469	0.9145	0.9747	0.9829	0.9803	0.9823
One-stage identification (YW)	0.0091	3.3726	0.8748	0.4734	0.5018	0.4517	0.4782	0.4522	0.4899	0.9475	0.9178	0.9679	0.9980	0.9823	0.9969
All-INAR(1) (YW)	0.0136	3.2319	0.8636	0.5171	0.5551	0.4842	0.5211	0.4839	0.5295	0.9471	0.9175	0.9765	0.9868	0.9826	0.9857
All-INARMA(1,1) (CLS)	-0.0461	3.4038	0.8869	0.5063	0.5401	0.4777	0.5125	0.4767	0.5209	0.9893	0.9536	1.0231	1.0293	1.0289	1.0276

Table 9-12 The effect of identification on INARMA forecasts for issue points (3000 series)

The identification method	ME	MSE	MASE	PB of MASE (INARMA/Benchmark)						RGRMSE (INARMA/Benchmark)					
				Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	Crost 0.2	Crost 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5
Two-stage identification (CLS)	0.0099	3.4099	0.9013	0.5202	0.5544	0.4888	0.5225	0.4883	0.5256	0.9907	0.9560	1.0076	1.0214	1.0160	1.0277
Two-stage identification (YW)	-0.0303	3.3835	0.8995	0.5190	0.5501	0.4827	0.5191	0.4814	0.5225	0.9653	0.9365	0.9879	1.0004	0.9973	1.0039
One-stage identification (CLS)	-0.0490	3.4327	0.9107	0.5107	0.5486	0.4783	0.5127	0.4780	0.5174	0.9961	0.9618	1.0188	1.0301	1.0282	1.0371
One-stage identification (YW)	-0.0572	3.5304	0.9202	0.5005	0.5408	0.4658	0.5144	0.4670	0.5161	0.9966	0.9749	1.0217	1.0469	1.0486	1.0527
All-INAR(1) (YW)	-0.0650	3.3563	0.9054	0.5117	0.5539	0.4756	0.5127	0.4744	0.5161	0.9927	0.9618	1.0200	1.0311	1.0293	1.0377
All-INARMA(1,1) (CLS)	-0.1366	3.5569	0.9327	0.5006	0.5378	0.4691	0.5027	0.4677	0.5097	1.0343	0.9940	1.0629	1.0707	1.0712	1.0772

The results of Table 9-11 and Table 9-12 also show that the two-stage identification procedure provides better results than the one-stage method. This agrees with the corresponding simulation results of section 8.6.2.3 for each of INARMA(0,0), INAR(1), INMA(1) and INARMA(1,1) models with similar parameters to those estimated for 3,000 series (see section 9.5.4 to find the corresponding parameters).

The CLS-based two-stage identification results for 3,000 series show that out of 1,943 filtered series, 54.55 percent were identified as INARMA(0,0), 23.88 percent as INAR(1), 17.96 percent as INMA(1), and 3.60 percent were identified as INARMA(1,1).

Now, if we go back to the results of the previous section, the results show that all-INARMA(1,1) results are worse than the benchmark methods. The simulation results also show that for the INAR(1) and INMA(1) processes with the corresponding parameters (the most common INARMA parameters of the 1,943 series are provided in the next section), the INARMA forecast accuracy is very close or slightly worse than that of benchmark methods. This could explain the 4 percent superiority (in terms of MSE) of the best benchmark forecasts (SBA 0.2) over the all-INARMA(1,1) forecasts.

Based on the results of this section, identification (two-stage) results in better INARMA forecasts than using an all-INAR(1) or all-INARMA(1,1) approach but the accuracy benefits are small. As also mentioned in the simulation chapter, the all-INAR(1) and all-INARMA(1,1) approaches are especially useful when the autoregressive parameter is high, but this is not the case for our empirical data.

9.5.4 INARMA(0,0), INAR(1), INMA(1) and INARMA(1,1) Series

As previously mentioned, a considerable percentage of series among 1,943 filtered series (of 3,000 series) were identified as INARMA(0,0), INAR(1) or INMA(1) and a few series were identified as INARMA(1,1). In this section, we separate the series according to the models identified and study them individually.

Because most of the series in the 16,000 data set are identified as INARMA(0,0) and a small number identified as other models, we do not present the results of each model here. Similar results for 16,000 series can be found in Appendix 9.A.

As the results of identification in the previous section suggest, 54.55 percent of 1943 series are identified as INARMA(0,0). Now, we assume that the order of these 1,060 series is taken to be (0,0) and the INARMA(0,0) forecasting method, i.e. using the average of all the previous observations as the forecast for the next period, is used. The results for all points in time and issue points are presented in Table 9-13 and Table 9-14.

Table 9-13 Only INARMA(0,0) series for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0288	-0.0843	0.1719	0.3017	0.1878	0.4303	0.1420
MSE	2.0878	2.3629	2.0998	2.2830	2.1043	2.3483	2.0749
MASE	0.9139	0.9819	0.9021	0.9352	0.9016	0.9382	0.8978
PB of MASE (INARMA/Benchmark)	0.5329	0.5624	0.5080	0.5453	0.5085	0.5491	-
RGRMSE (INARMA/Benchmark)	0.8992	0.8543	0.9111	0.8922	0.9152	0.8798	-

Table 9-14 Only INARMA(0,0) series for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0071	-0.0959	0.1482	0.2827	0.1639	0.4089	0.0830
MSE	2.1982	2.4911	2.2013	2.3901	2.2049	2.4482	2.1704
MASE	0.9472	1.0182	0.9327	0.9635	0.9318	0.9634	0.9298
PB of MASE (INARMA/Benchmark)	0.5328	0.5642	0.5023	0.5388	0.5013	0.5362	-
RGRMSE (INARMA/Benchmark)	0.9743	0.9149	0.9763	0.9549	0.9836	0.9450	-

Investigating the estimated parameter of the INARMA(0,0) process ($\hat{\lambda}$), we found that in general $\hat{\lambda}$ is close to 1 (the average is 1.2641 and 73.87 percent are between 0.5 and 1.5).

The results of Table 9-13 agree with the corresponding simulation results ($\lambda = 1$ and $n = 24$). The simulation results show that INARMA produces the best results and the empirical results also show a very narrow improvement over the best benchmark method. The PB of MASE results confirm the results of MASE for both all points in time and issue points.

The results of identification show that 23.88 percent of 1,943 series are identified as INAR(1). Now, we assume that the order of these 464 series is taken to be (1,0) and the INAR(1) forecasting method is used. We also investigate the effect of using different estimation methods on forecast accuracy. For this reason, the forecasts based on the three estimation methods of CLS, YW and CML are presented in Table 9-15 and Table 9-16.

In general, the estimated autoregressive parameter of the INAR(1) process, $\hat{\alpha}$, is close to 0.1 (the average is 0.1234 and 50.65 percent are between 0.05 and 0.15) and the estimated innovation parameter, $\hat{\lambda}$, is around 2 (the average is 2.5972 and 40.52 percent are between 1 and 3). The simulation results for the corresponding parameters (similar values of α, λ, and n) show that YW yields better results than CLS and CML. The simulation results also show that the YW results are slightly better than those of SBA and SBJ with smoothing parameter 0.2.

Table 9-15 Only INAR(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.1557	-0.1498	0.1365	0.5791	0.1689	0.8221	-0.1963	-0.1411	-0.1257
MSE	4.8743	5.6076	4.8235	5.5217	4.8319	5.8396	4.9049	4.7921	4.8386
MASE	0.9168	0.9721	0.8923	0.9218	0.8906	0.9312	0.9203	0.9088	0.9177
PB of MASE INAR(1)-CLS/Benchmark	0.4989	0.5408	0.4596	0.4919	0.4605	0.5068			
PB of MASE INAR(1)-YW/Benchmark	0.5119	0.5458	0.4682	0.5038	0.4662	0.5135			
PB of MASE INAR(1)-CML/Benchmark	0.5083	0.5462	0.4657	0.4928	0.4653	0.5070			
RGRMSE INAR(1)-CLS/Benchmark	0.9940	0.9733	1.0499	1.0924	1.0529	1.1180			
RGRMSE INAR(1)-YW/Benchmark	0.9726	0.9554	1.0275	1.0703	1.0311	1.0950			
RGRMSE INAR(1)-CML/Benchmark	0.9887	0.9672	1.0443	1.0835	1.0457	1.1087			

Table 9-16 Only INAR(1) series for issue points (3000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.1439	-0.1390	0.1465	0.5856	0.1787	0.8272	-0.2755	-0.2165	-0.2114
MSE	5.0350	5.7658	4.9933	5.7037	5.0027	6.0297	5.1167	4.9863	5.0296
MASE	0.9600	1.0166	0.9362	0.9669	0.9346	0.9764	0.9743	0.9609	0.9685
PB of MASE INAR(1)-CLS/Benchmark	0.4850	0.5330	0.4464	0.4837	0.4473	0.4962			
PB of MASE INAR(1)-YW/Benchmark	0.4988	0.5411	0.4579	0.4954	0.4551	0.5024			
PB of MASE INAR(1)-CML/Benchmark	0.5005	0.5413	0.4544	0.4872	0.4552	0.5001			
RGRMSE INAR(1)-CLS/Benchmark	1.0272	1.0087	1.0864	1.1242	1.0913	1.1743			
RGRMSE INAR(1)-YW/Benchmark	1.0041	0.9888	1.0625	1.1012	1.0682	1.1496			
RGRMSE INAR(1)-CML/Benchmark	1.0177	0.9981	1.0731	1.1107	1.0778	1.1582			

The results of Table 9-15 agree with the simulation results in that YW results are better than CLS and CML. The YW-based INAR(1) forecasts are also better than benchmarks by one percent in terms of MSE. However, the MASE of INAR(1) is worse than that of the best benchmark by one percent.

The results of Table 9-16 also show a narrow improvement of INAR(1)-YW compared to the best benchmark (0.15 percent in terms of MSE) although the INARMA forecasts are biased when only issue points are considered.

Based on the results of identification, 17.96 percent of 1,943 series are identified as INMA(1). We assume that the order of these 349 series is taken to be (0,1) and the INMA(1) forecasting method is used. We also investigate the effect of using different estimation methods (CLS and YW) on the forecasting accuracy. The results are given in Table 9-17 and Table 9-18.

Table 9-17 Only INMA(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	-0.0869	-0.0875	0.1907	0.6066	0.2215	0.8380	-0.1041	-0.0936
MSE	4.3556	4.9339	4.3470	4.9673	4.3581	5.2745	4.3291	4.3464
MASE	0.6681	0.7022	0.6556	0.6778	0.6549	0.6904	0.6676	0.6668
PB of MASE INMA(1)-CLS/Benchmark	0.5093	0.5418	0.4728	0.5033	0.4702	0.5150		
PB of MASE INMA(1)-YW/Benchmark	0.5150	0.5461	0.4742	0.5062	0.4723	0.5177		
RGRMSE INMA(1)-CLS/Benchmark	0.9546	0.9687	0.9829	1.0729	0.9960	1.0516		
RGRMSE INMA(1)-YW/Benchmark	0.9345	0.9527	0.9642	1.0524	0.9771	1.0316		

Table 9-18 Only INMA(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	-0.0870	-0.0846	0.1889	0.6043	0.2195	0.8340	-0.1593	-0.1444
MSE	4.4429	5.0046	4.4358	5.0503	4.4471	5.3611	4.4548	4.4742
MASE	0.7177	0.7512	0.7045	0.7255	0.7039	0.7391	0.7256	0.7235
PB of MASE INMA(1)-CLS/Benchmark	0.4924	0.5304	0.4575	0.4891	0.4539	0.5019		
PB of MASE INMA(1)-YW/Benchmark	0.5026	0.5361	0.4616	0.4945	0.4594	0.5064		
RGRMSE INMA(1)-CLS/Benchmark	0.9891	1.0325	1.0230	1.1233	1.0399	1.1033		
RGRMSE INMA(1)-YW/Benchmark	0.9654	1.0118	0.9987	1.0970	1.0152	1.0773		

Looking at the estimated parameters of the INMA(1) process $(\hat{\beta}, \hat{\lambda})$ reveals that in general, $\hat{\beta}$ is close to zero (the average is 0.0374 and 79.94 percent are between 0 and 0.05) and $\hat{\lambda}$ is between 2 and 3 (the average is 2.7357 and 43.55 percent are between 2 and 3).

The simulation results for the corresponding parameters (similar values of β, λ, and n) show that YW yields slightly better results than CLS and also both CLS and YW results are close to SBA and SBJ with smoothing parameter 0.2 (with INMA slightly better in terms of MSE). The results of Table 9-17 show that YW forecasts are better than CLS in terms of MASE (only by 0.1 percent) but worse in terms of MSE (only by 0.5 percent). The best INMA(1) is better than the best benchmark by only 0.4 percent using MSE and worse than the best benchmark by 1.8 percent using MASE.

Finally, the results of identification show that 3.60 percent of 1,943 series are identified as INARMA(1,1). We assume that the order of these 70 series is taken to be (1,1) and the INARMA(1,1) forecasting method is used. We also investigate the effect of using different estimation methods (CLS and YW) on the forecasting accuracy. The results are given in Table 9-19 and Table 9-20.

Table 9-19 Only INARMA(1,1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	-0.1340	-0.1141	0.1556	0.6049	0.1878	0.8446	-0.2016	-0.0997
MSE	4.7747	5.3016	4.7204	5.2908	4.7275	5.6017	5.0449	5.8052
MASE	0.8853	0.9369	0.8695	0.9076	0.8690	0.9192	0.9178	0.9849
PB of MASE INMA(1)-CLS/Benchmark	0.4607	0.5226	0.4452	0.4821	0.4512	0.5036		
PB of MASE INMA(1)-YW/Benchmark	0.4417	0.4881	0.4369	0.4702	0.4429	0.4905		
RGRMSE INMA(1)-CLS/Benchmark	1.0854	1.0284	1.1681	1.0737	1.1793	1.1495		
RGRMSE INMA(1)-YW/Benchmark	0.8169	0.7752	0.8747	0.7943	0.8910	0.8146		

In general, the estimated autoregressive parameter of an INARMA(1,1) process is in the range $0.1 < \hat{\alpha} < 0.3$ (the average is 0.1907 and 54.28 percent are between 0.05 and 0.35), the moving average parameter, $\hat{\beta}$, is close to zero (the average is 0.0773 and 57.14 percent are between 0.01 and 0.1) and the innovation parameter, $\hat{\lambda}$, is around 2 (the average is 2.1996 and 67.14 percent are between 1 and 2.5).

Table 9-20 Only INARMA(1,1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	-0.1585	-0.1414	0.1291	0.5734	0.1610	0.8116	-0.3483	-0.2846
MSE	5.0255	5.6064	4.9530	5.5318	4.9579	5.8222	5.2348	5.8600
MASE	0.9274	0.9841	0.9112	0.9482	0.9106	0.9573	0.9576	1.0133
PB of MASE INMA(1)-CLS/Benchmark	0.4564	0.5269	0.4426	0.4849	0.4509	0.5019		
PB of MASE INMA(1)-YW/Benchmark	0.4521	0.5029	0.4390	0.4812	0.4458	0.4946		
RGRMSE INMA(1)-CLS/Benchmark	1.0896	1.0261	1.1812	1.1056	1.1952	1.2051		
RGRMSE INMA(1)-YW/Benchmark	0.8444	0.7778	0.8989	0.8381	0.9221	0.8584		

The simulation results for corresponding parameters (similar values of α, β, λ, and n) show that CLS produces better forecasts than YW. The simulation results also show that SBA and SBJ have better results in terms of MSE and MASE by up to one percent. The results of Table 9-19 show that the MSE of the best benchmark method (SBA 0.2) is better than that of INARMA(1,1)-CLS by 6.5 percent and by 5.3 percent for MASE.

The results of Table 9-20 also show that CLS produces better forecasts using MSE and MASE when only issue points are considered, agreeing with simulation results (not presented). The best benchmark method is still better than INARMA(1,1)-CLS by 5.38 percent in terms of MSE and 4.90 percent in terms of MASE.

9.5.5 *h*-step-ahead Forecasts for INAR(1) Series

All of the forecasts in section 9.5.4 were one-step-ahead. In this subsection, the results of three-step and six-step-ahead forecasts for INAR(1) series of 16,000 and 3,000 series are presented. Here, we only use YW-based forecasts because of their better performance for INAR(1) processes compared to CLS and CML. This has been shown both for theoretically generated data (section 8.3) and empirical data (sections 9.5.1 and 9.5.4).

The simulation results show that, for INAR(1) processes, the performance of INARMA over the benchmark methods is improved for a *h*-step ahead compared to the one-step ahead case when the autoregressive parameter is low. But when the autoregressive parameter is high, the forecasts converge to the mean of process, resulting in poor INARMA forecasts. The simulation results also show that when more observations are available, the improvement by using INARMA over the benchmark methods increases.

Because the autoregressive parameter of the 16,000 series is estimated to be low (see appendix 9.A), we see a substantial improvement over the best benchmark by using an INAR(1) forecast. For $h = 3$, the results of Table 9-21 show that the improvement is 9.2 percent using MSE and 16.4 percent using MASE compared to the best benchmark. It can be seen that the improvement in terms of MSE is more than that of the one-step ahead forecasts (which is 7.2 percent), agreeing with the simulation results. However, the improvement in terms of MASE is less than that of the one-step ahead forecasts (which is 17.7 percent).

Table 9-21 Three-step-ahead YW-INAR(1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.1237	-0.1686	-0.0583	0.0061	-0.0510	0.0643	0.0487
MSE	0.9643	1.0393	0.9355	0.9330	0.9332	0.9236	0.8475
MASE	2.2896	2.3359	2.1679	2.0109	2.1545	1.9114	1.5977
PB of MASE INAR(1)/Benchmark	0.5943	0.5493	0.5666	0.5095	0.5597	0.4706	-
RGRMSE INAR(1)/Benchmark	0.8654	0.9516	0.9060	1.1389	0.9119	1.2063	-

Table 9-22 Three-step-ahead YW-INAR(1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.0389	-0.0773	0.0152	0.0678	0.0213	0.1162	0.0193
MSE	0.9931	1.0865	0.9915	1.0363	0.9922	1.0447	1.0159
MASE	0.6814	0.6774	0.6575	0.6143	0.6550	0.5976	0.6560
PB of MASE INAR(1)/Benchmark	0.5533	0.4839	0.5192	0.3790	0.5069	0.3485	-
RGRMSE INAR(1)/Benchmark	1.0231	1.3512	1.0843	1.6532	1.0916	1.7859	-

The results of Table 9-23 show that, when $h = 6$, we also see a substantial improvement in using an INAR(1) forecast. The improvement is 7.8 percent using

MSE and 16.2 percent using MASE compared to the best benchmark. It can be seen that the improvement using MSE is slightly more than that of one-step ahead forecast. However, the improvement using MASE again deteriorates by 1.5 percent compared to the one-step ahead forecast.

Table 9-23 Six-step-ahead YW-INAR(1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.1203	-0.1612	-0.0545	0.0136	-0.0472	0.0718	0.0576
MSE	1.0029	1.0706	0.9738	0.9656	0.9715	0.9567	0.8822
MASE	2.3385	2.3866	2.2119	2.0483	2.1979	1.9453	1.6286
PB of MASE INAR(1)/Benchmark	0.6139	0.5503	0.5731	0.4810	0.5664	0.4573	-
RGRMSE INAR(1)/Benchmark	0.8534	0.9415	0.8999	1.1492	0.9055	1.2119	-

Table 9-24 Six-step-ahead YW-INAR(1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.0818	-0.1170	-0.0282	0.0257	-0.0223	0.0733	-0.0288
MSE	0.9266	1.0096	0.9189	0.9449	0.9188	0.9480	0.9428
MASE	0.6656	0.6651	0.6387	0.5937	0.6357	0.5731	0.6358
PB of MASE INAR(1)/Benchmark	0.5653	0.4500	0.5359	0.3580	0.5173	0.3530	-
RGRMSE INAR(1)/Benchmark	0.9913	1.3136	1.0633	1.6153	1.0709	1.7608	-

As discussed in the previous section, the autoregressive parameter of the 3,000 series is estimated to be close to zero. Therefore, because the estimated $\hat{\alpha}$ is low, we see an improvement by using INAR(1) over the benchmark methods. The improvement is 1.8 percent using MSE. But the MASE is worse than the best benchmark by 1.7 percent. It can be seen that the improvement in terms of MSE is more than that of the one-step ahead forecasts (0.6 percent). Also the deterioration in terms of MASE has been decreased compared to one-step ahead case (which was 2 percent), agreeing with the simulation results.

Table 9-25 Three-step-ahead YW-INAR(1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.2430	-0.2309	0.0516	0.5026	0.0844	0.7471	-0.2403
MSE	4.9832	5.8989	4.8563	5.5920	4.8567	5.8457	4.7689
MASE	0.9188	0.9835	0.8885	0.9188	0.8862	0.9234	0.9020
PB of MASE INAR(1)/Benchmark	0.5276	0.5459	0.4735	0.5017	0.4698	0.5022	-
RGRMSE INAR(1)/Benchmark	0.9893	0.9866	1.0603	1.1242	1.0638	1.1425	-

Table 9-26 Three-step-ahead YW-INAR(1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.2364	-0.2219	0.0561	0.5058	0.0886	0.7483	-0.2654
MSE	5.0839	5.9960	4.9608	5.7035	4.9616	5.9623	4.8882
MASE	0.9519	1.0173	0.9211	0.9496	0.9187	0.9541	0.9390
PB of MASE INAR(1)/Benchmark	0.5227	0.5471	0.4697	0.4975	0.4641	0.4976	-
RGRMSE INAR(1)/Benchmark	1.0096	1.0098	1.0880	1.1553	1.0874	1.1650	-

The results of the six-step ahead compared to one-step ahead forecasts show that the improvement using MSE (0.2 percent) is decreased. The deterioration in terms of MASE (2.3 percent) is also increased.

The corresponding simulation results also show that when we compare the three-step and one-step ahead forecasts, there is an improvement in terms of MSE (1.8 percent) and a deterioration in terms of MASE (1.7 percent) is decreased. But for the six-step ahead case, the deterioration using MASE is increased compared to one-step ahead case (2.3 percent vs. 2 percent). Therefore, the only difference between the empirical and simulation results is regarding the MSE of six-step ahead forecasts, which has not been improved compared to the one-step ahead forecasts.

The results of three-step and six-step forecasts for INARMA(0,0), INMA(1) and INARMA(1,1) for both 16,000 and 3,000 series have also been found. Since these results also agree with INAR(1) results, we do not present them here (see Appendix 9.B). In general, the performance of INARMA in relation to the benchmarks improves for *h*-step ahead compared to one-step ahead forecasts.

Table 9-27 Six-step-ahead YW-INAR(1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.3948	-0.4292	-0.0939	0.3318	-0.0604	0.5855	-0.3495
MSE	5.3005	6.5011	5.0488	5.7594	5.0366	5.9095	5.0277
MASE	0.9254	1.0090	0.8865	0.9179	0.8832	0.9130	0.9041
PB of MASE INAR(1)/Benchmark	0.5194	0.5640	0.4667	0.4994	0.4603	0.4865	-
RGRMSE INAR(1)/Benchmark	1.0012	0.9946	1.0789	1.1489	1.0845	1.1834	-

Table 9-28 Six-step-ahead YW-INAR(1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1)
ME	-0.4144	-0.4415	-0.1157	0.3120	-0.0825	0.5632	-0.4031
MSE	5.3790	6.6227	5.1032	5.8224	5.0885	5.9554	5.1182
MASE	0.9566	1.0448	0.9147	0.9451	0.9112	0.9381	0.9393
PB of MASE INAR(1)/Benchmark	0.5107	0.5585	0.4594	0.4928	0.4532	0.4810	-
RGRMSE INAR(1)/Benchmark	1.0245	1.0320	1.1112	1.1790	1.1168	1.2345	-

9.5.6 The Effect of Length of History

It has been shown in chapter 8 that when more observations are available, the accuracy of INARMA forecasts improves at a greater rate than benchmark methods. The effect of length of history on the accuracy of forecasts produced by different methods is examined in this section. In order to do so, we need a data set with more than 24 periods; therefore we are able to use only the 16,000 series. The last 12 periods have been assigned as the performance period and the length of estimation period changes from 12 to 60 periods. This is shown in Table 9-29.

Table 9-29 The estimation and performance periods

Cases	Estimation period	Performance period
Case 1	49-60	61-72
Case 2	13-60	61-72
Case 3	1-60	61-72

The INARMA results are based on all-INAR(1) forecasts with YW estimates. The results for each of the above cases are presented in Table 9-30 to Table 9-35.

Table 9-30 The forecasting accuracy for all points in time for case 1

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2043	-0.2139	-0.1501	-0.0760	-0.1441	-0.0300	-0.0225
MSE	0.6827	0.7154	0.6408	0.6157	0.6369	0.6027	0.5201
MASE	1.9684	1.9757	1.8422	1.6583	1.8283	1.5553	1.2098
PB of MASE (INARMA/Benchmark)	0.6364	0.6426	0.5951	0.5328	0.5912	0.4969	-
RGRMSE (INARMA/Benchmark)	0.8312	0.7989	0.8715	0.9580	0.8784	1.0404	-

Table 9-31 The forecasting accuracy for issue points for case 1

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1664	-0.1652	-0.1163	-0.0403	-0.1108	0.0014	-0.1321
MSE	1.0008	1.0671	0.9692	0.9866	0.9668	0.9887	0.9688
MASE	0.4875	0.4858	0.4646	0.4290	0.4621	0.4122	0.4738
PB of MASE (INARMA/Benchmark)	0.4921	0.4517	0.4300	0.3440	0.4232	0.3207	-
RGRMSE (INARMA/Benchmark)	1.0452	1.0694	1.1223	1.3284	1.1313	1.4529	-

Table 9-32 The forecasting accuracy for all points in time for case 2

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0709	-0.0948	-0.0381	-0.0070	-0.0345	0.0223	0.0258
MSE	0.4257	0.4616	0.4146	0.4244	0.4138	0.4237	0.3867
MASE	2.2628	2.2784	2.1383	1.9623	2.1245	1.8590	1.5077
PB of MASE (INARMA/Benchmark)	0.6714	0.6761	0.6154	0.5377	0.6086	0.4873	-
RGRMSE (INARMA/Benchmark)	0.8032	0.7836	0.8632	0.9692	0.8716	1.0617	-

Table 9-33 The forecasting accuracy for issue points for case 2

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0417	-0.0626	-0.0073	0.0288	-0.0034	0.0592	-0.0552
MSE	0.8668	0.9500	0.8574	0.9024	0.8572	0.9097	0.8655
MASE	0.3620	0.3713	0.3482	0.3335	0.3467	0.3226	0.3820
PB of MASE (INARMA/Benchmark)	0.4452	0.4417	0.3827	0.3158	0.3742	0.2835	-
RGRMSE (INARMA/Benchmark)	1.2041	1.2695	1.3042	1.5976	1.3173	1.7591	-

Table 9-34 The forecasting accuracy for all points in time for case 3

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0523	-0.0788	-0.0218	0.0041	-0.0184	0.0317	0.0330
MSE	0.4057	0.4415	0.3980	0.4117	0.3976	0.4127	0.3838
MASE	2.2626	2.2831	2.1493	1.9939	2.1368	1.8996	1.6285
PB of MASE (INARMA/Benchmark)	0.6816	0.6874	0.6243	0.5507	0.6176	0.4966	-
RGRMSE (INARMA/Benchmark)	0.7986	0.7771	0.8592	0.9630	0.8675	1.0552	-

Table 9-35 The forecasting accuracy for issue points for case 3

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0313	-0.0560	0.0021	0.0337	0.0058	0.0636	-0.0464
MSE	0.8615	0.9447	0.8531	0.8985	0.8529	0.9060	0.8621
MASE	0.3519	0.3636	0.3389	0.3274	0.3375	0.3170	0.3743
PB of MASE (INARMA/Benchmark)	0.4436	0.4571	0.3787	0.3269	0.3717	0.2889	-
RGRMSE (INARMA/Benchmark)	1.2297	1.2857	1.3336	1.6196	1.3467	1.7840	-

The results confirm that when the length of estimation period is longer, the accuracy of all methods in terms of MSE improves. The PB of MASE and RGRMSE also improve when more data are available.

The results show that when more observations become available, the MASE of all methods increases. However, the PB of MASE increases when the length of history increases. As argued in section 9.5.1, the results of MASE are not reliable for highly intermittent data and, therefore, we will not pursue these results.

9.5.7 Lead Time Forecasting for INAR(1) Series

The conditional expected values of the over-lead-time-aggregated INARMA series were obtained in chapter 6. The results for the four INARMA processes are summarized in Equation 9-7.

$$E\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = \begin{cases} (l+1)\lambda & \text{for INARMA(0,0)} \\ \dfrac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \dfrac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right] & \text{for INAR(1)} \\ (l+1)(1+\beta)\lambda & \text{for INMA(1)} \\ \dfrac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \dfrac{\lambda(1+\beta)}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right] & \text{for INARMA(1,1)} \end{cases}$$

Equation 9-7

The accuracy of INARMA lead time forecasts has been compared to that of benchmark methods. This has been done for each INARMA model of section 9.5.4, individually. Based on the previous findings, the parameters of INAR(1) and INMA(1) processes are estimated using YW and those of an INARMA(1,1) process are estimated by CLS.

The lead time forecast for benchmark methods is simply the length of lead time multiplied by the one-step-ahead forecast. Two values for lead time have been considered: $l = 3, 6$.

In order to test the benefit by using the INARMA lead time forecasts (hereafter abbreviated as INARMA-LT), these results have been compared to the results of using cumulative h-step ahead forecasts over lead time (hereafter abbreviated as INARMA-h). The latter is given by $\sum_{i=1}^{l+1} \hat{Y}_{t+i}$, where $\hat{Y}_{t+i}$ is the i-step ahead forecast.

The results of lead time forecasting for INAR(1) series of both 16,000 and 3,000 series are presented in this section. The results for INARMA(0,0), INMA(1) and INARMA(1,1) series are provided in Appendix 9.C.

The results of Table 9-36 for INAR(1) series of 16,000 series show that when $l = 3$, INAR(1)-LT improves the MSE by 22.8 percent and MASE by 16.1 percent compared to the best benchmark (which is SBJ 0.5). It can be seen that the results of INAR(1)-h are worse than INAR(1)-LT. However, these results still improve the MSE by 19.8 percent and the MASE by 14.9 percent compared to the best benchmark.

Table 9-36 Lead-time forecasts $(l = 3)$ for INAR(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.3884	-0.5232	-0.1922	0.0010	-0.1704	0.1757	0.1159	0.1304
MSE	3.9820	4.7702	3.7116	3.7566	3.6895	3.6534	2.8205	2.9269
MASE	2.9812	3.1303	2.8159	2.6880	2.8002	2.5686	2.1538	2.1860
PB of MASE INARMA-LT/Benchmark	0.5597	0.5779	0.5389	0.5545	0.5381	0.5467		
PB of MASE INARMA-h/Benchmark	0.5337	0.5839	0.5260	0.5692	0.5234	0.5398		
RGRMSE INARMA-LT/Benchmark	0.9078	0.8863	0.9913	0.9647	0.9642	0.9824		
RGRMSE INARMA-h/Benchmark	0.9574	0.9345	1.0452	1.0139	1.0141	1.0352		

Table 9-37 Lead-time forecasts $(l = 3)$ for INAR(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.0871	-0.2020	0.0755	0.2332	0.0936	0.3783	-0.3597	0.1188
MSE	3.3138	4.1399	3.3198	3.7409	3.3277	3.8344	3.6080	3.5530
MASE	1.0419	1.0990	1.0122	1.0145	1.0103	1.0022	1.2490	1.0374
PB of MASE INARMA-LT/Benchmark	0.3662	0.4498	0.3512	0.3813	0.3473	0.3851		
PB of MASE INARMA-h/Benchmark	0.5130	0.5206	0.5140	0.4813	0.5068	0.4395		
RGRMSE INARMA-LT/Benchmark	1.4421	1.9523	1.5540	2.3331	1.5482	2.5471		
RGRMSE INARMA-h/Benchmark	1.0805	1.2766	1.1410	1.4439	1.1287	1.5435		

Table 9-38 Lead-time forecasts $(l = 6)$ for INAR(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.8168	-1.0621	-0.4219	-0.0134	-0.3780	0.3362	0.2380	0.2553
MSE	12.5629	15.6078	11.3849	11.3478	11.2858	10.8684	7.2081	7.9582
MASE	4.7715	5.0536	4.4861	4.2692	4.4574	4.0743	3.1157	3.2439
PB of MASE INARMA-LT/Benchmark	0.5901	0.5873	0.5844	0.5996	0.5825	0.5930		
PB of MASE INARMA-h/Benchmark	0.5503	0.5901	0.5427	0.6034	0.5465	0.5721		
RGRMSE INARMA-LT/Benchmark	0.9137	0.8081	0.8685	0.9215	0.8673	0.9763		
RGRMSE INARMA-h/Benchmark	0.9746	0.8833	0.9340	1.0050	0.9331	1.0669		

When $l = 6$, INAR(1)-LT improves the MSE by 33.7 percent and the MASE by 23.5

percent compared to the best benchmark (which is SBJ 0.5). Again, the INAR(1)-h results are worse than those of INAR(1)-LT. But, these results still improve the MSE by 26.7 percent and the MASE by 20.4 percent compared to the best benchmark.

Table 9-39 Lead-time forecasts $(l = 6)$ for INAR(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.1007	-0.3119	0.2208	0.5447	0.2566	0.8302	-0.2983	0.2971
MSE	8.3539	11.4468	8.4132	9.9329	8.4479	10.3190	8.8087	9.7652
MASE	1.5914	1.7284	1.5548	1.5932	1.5524	1.5935	1.7465	1.6419
PB of MASE INARMA-LT/Benchmark	0.4394	0.4688	0.4617	0.4303	0.4485	0.4475		
PB of MASE INARMA-h/Benchmark	0.4776	0.5107	0.4886	0.5087	0.4972	0.4975		
RGRMSE INARMA-LT/Benchmark	1.4130	1.6212	1.3792	1.9236	1.4565	2.1010		
RGRMSE INARMA-h/Benchmark	1.2483	1.2921	1.2192	1.4892	1.3122	1.6324		

According to the results of Table 9-39, when only issue points are considered, the INARMA results are again worse than the best benchmark method. In this case, the INAR(1)-h produces better forecasts than INAR(1)-LT in terms of MASE but the MSE of INAR(1)-LT is smaller.

Table 9-40 Lead-time forecasts $(l = 3)$ for INAR(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.6244	-0.5881	0.2594	1.6123	0.3576	2.3458	-0.5794	-0.6386
MSE	19.5567	27.2937	18.6953	25.3595	18.7301	27.9187	17.8177	18.3578
MASE	1.5641	1.8153	1.5005	1.6966	1.4973	1.7523	1.5021	1.5179
PB of MASE INARMA-LT/Benchmark	0.5416	0.6002	0.4866	0.5491	0.4832	0.5575		
PB of MASE INARMA-h/Benchmark	0.5241	0.5737	0.4800	0.5394	0.4772	0.5556		
RGRMSE INARMA-LT/Benchmark	0.9835	0.9097	1.0562	1.0638	1.0655	1.0872		
RGRMSE INARMA-h/Benchmark	1.0312	0.9614	1.1130	1.1243	1.1232	1.1483		

The results for INAR(1) series of 3,000 series show that when $l = 3$, INAR(1)-LT improves the MSE of forecast by 4.7 percent compared to the best benchmark (which is SBA 0.2). The same is not true for MASE which is 0.32 percent worse than that of SBJ 0.2. The results of INAR(1)-h are worse than those of INAR(1)-LT. But, these results still improve the MSE by 1.8 percent. The MASE of INAR(1)-h is worse than the best benchmark by 1.3 percent.

Table 9-41 Lead-time forecasts $(l = 3)$ for INAR(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.6199	-0.5765	0.2576	1.6065	0.3551	2.3342	-0.7597	-0.6880
MSE	20.0611	27.9048	19.1931	25.9631	19.2274	28.5220	18.5668	18.9400
MASE	1.6365	1.8978	1.5715	1.7717	1.5683	1.8269	1.5926	1.5973
PB of MASE INARMA-LT/Benchmark	0.5316	0.5891	0.4704	0.5417	0.4653	0.5481		
PB of MASE INARMA-h/Benchmark	0.5169	0.5739	0.4750	0.5361	0.4725	0.5511		
RGRMSE INARMA-LT/Benchmark	1.0259	0.9565	1.0988	1.1158	1.1065	1.1248		
RGRMSE INARMA-h/Benchmark	1.0557	0.9906	1.1326	1.1485	1.1421	1.1604		

Based on the results of Table 9-41, when only issue points are considered, the INARMA results are still better than the best benchmark method in terms of MSE. Again, the INAR(1)-LT produces better forecasts than INAR(1)-h in terms of MSE and MASE, although the benefit is slight.

It can be seen from Table 9-42 that when $l = 6$, INAR(1)-LT improves the MSE of forecast by 3.6 percent compared to the best benchmark (which is SBJ 0.2). However, MASE is 3.3 percent worse than that of SBJ 0.2. The results of INAR(1)-h are worse than those of INAR(1)-LT. These results are worse than the best benchmark by 4.1 percent in terms of MSE and 5.9 percent in terms of MASE.

Table 9-42 Lead-time forecasts $(l = 6)$ for INAR(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.9213	-2.1274	-0.1155	2.4386	0.0852	3.9606	-1.6259	-1.6279
MSE	55.4939	93.3985	48.8035	73.9507	48.6269	81.7715	46.8761	50.7215
MASE	2.4988	3.1026	2.2948	2.7640	2.2829	2.8890	2.3615	2.4279
PB of MASE INARMA-LT/Benchmark	0.5551	0.6192	0.4751	0.5579	0.4695	0.5665		
PB of MASE INARMA-h/Benchmark	0.5179	0.5887	0.4671	0.5493	0.4655	0.5591		
RGRMSE INARMA-LT/Benchmark	1.0384	0.9309	1.1841	1.1827	1.1810	1.1967		
RGRMSE INARMA-h/Benchmark	1.1664	1.0574	1.3307	1.3041	1.3402	1.2950		

Table 9-43 Lead-time forecasts $(l = 6)$ for INAR(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.9801	-2.1428	-0.1878	2.3786	0.0113	3.8857	-2.0011	-1.7986
MSE	58.5370	97.4990	51.2152	76.6573	50.9742	84.1030	50.6760	53.9097
MASE	2.6665	3.2866	2.4389	2.9059	2.4247	3.0262	2.5553	2.6023
PB of MASE INARMA-LT/Benchmark	0.5419	0.6074	0.4578	0.5401	0.4499	0.5457		
PB of MASE INARMA-h/Benchmark	0.5148	0.5840	0.4629	0.5383	0.4586	0.5447		
RGRMSE INARMA-LT/Benchmark	1.0680	0.9990	1.2407	1.3281	1.2650	1.3375		
RGRMSE INARMA-h/Benchmark	1.1846	1.1087	1.3696	1.4120	1.4009	1.3771		

According to the results of Table 9-43, when only issue points are considered, the INARMA results are still slightly better than the best benchmark method in terms of MSE (only 0.58 percent). The INAR(1)-LT results are again better than those of INAR(1)-h in terms of MSE and MASE.

In general, the results show that the difference between the lead time forecasts for INAR(1)-LT and the benchmarks is greater than that of one-step ahead forecasts. Although the MASE of INAR(1) is worse than the best benchmark (only for 3,000 series), the difference is less than that of one-step ahead. This is true for both 3,000 and 16,000 series, although the improvement is greater for the latter. This could suggest that with a longer length of history, INARMA performance improves with a greater rate compared to the benchmarks.

Comparing the results of using the lead time forecasts of Equation 9-7 with the results of using the cumulative *h*-step ahead forecasts shows that the former produces better forecasts using MSE and MASE.

The results of Appendix 9.C for INARMA(0,0) and INMA(1) series are similar to the above results for INAR(1) series. This means that for both processes the INARMA method produces more accurate lead time forecasts than benchmarks. For INMA(1) series, the INMA(1)-LT results are better than those of INMA(1)-h. However, the same is not true for INARMA(1,1) series. For these series, INARMA(1,1)-h beats INARMA(1,1)-LT forecasts and only the former is better than the best benchmark. It should be noted that only a few series were identified as

INARMA(1,1) for both 16,000 and 3,000 series. Also because it has been shown both in simulation and empirical analysis that an all-INAR(1) approach is promising, we will not pursue the poor performance of INARMA-LT compared to INARMA-h for INARMA(1,1) series and leave it for further research.

The following section focuses on comparing the lead time forecasts of the all-INAR(1) method and benchmark methods for all 16,000 and 3,000 series.

9.5.8 Lead Time Forecasting for all-INAR(1)

The previous section only focused on those series identified as INAR(1). In this section, an all-INAR(1) approach is used for all 16,000 and 3,000 series (5,168 and 1,943 filtered series).

According to the results of section 9.5.3, the forecasts produced by the all-INAR(1) approach are close to those based on identification. Here we compare the accuracy of lead time forecasts of all-INAR(1) and benchmarks for $l = 3, 6$.

Again, the lead time forecasts of Equation 9-7 will be compared to results of using the cumulative h-step ahead forecasts over lead time.

The results for 16,000 series show that for $l = 3$, when an all-INAR(1) approach is used, the INARMA-LT forecasts are better than the best benchmark by 18.7 percent using MSE and 13.5 percent using MASE. Although INARMA-h forecasts are not as good as INARMA-LT forecasts, they still outperform the best benchmark by 17.1 percent using MSE and 13.1 using MASE.

The results of Table 9-45 show that when only issue points are considered, the INAR(1)-h results are better than those of INAR(1)-LT in terms of both MSE and MASE. The INAR(1)-h forecasts have also smaller MSE and MASE than the best benchmark method.

Comparing these results with the results of the previous section shows that INARMA-LT and INARMA-h forecasts are close when all the data series (and not only INAR(1) series as in the previous section) are considered. This is because, as discussed in section 9.5.3, the majority of the 16,000 series are INARMA(0,0) and, for them, the two methods are the same.

Table 9-44 Lead-time forecasts $(l = 3)$ for all-INAR(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.2283	-0.2909	-0.1326	-0.0359	-0.1219	0.0491	0.0840	0.0958
MSE	1.6254	1.8912	1.5207	1.5625	1.5126	1.5458	1.2287	1.2534
MASE	5.5815	5.6849	5.2803	4.9029	5.2476	4.6645	4.0331	4.0511
PB of MASE INARMA-LT/Benchmark	0.6145	0.6338	0.5831	0.5450	0.5790	0.5143		
PB of MASE INARMA-h/Benchmark	0.6121	0.6335	0.5793	0.5411	0.5751	0.5081		
RGRMSE INARMA-LT/Benchmark	7.0337	0.8043	0.8923	0.9406	0.8960	1.0031		
RGRMSE INARMA-h/Benchmark	10.4127	0.8161	0.9049	0.9550	0.9089	1.0159		

Table 9-45 Lead-time forecasts $(l = 3)$ for all-INAR(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.1423	-0.1925	-0.0535	0.0419	-0.0437	0.1201	-0.0426	0.0971
MSE	2.0187	2.3511	1.9216	2.0165	1.9149	2.0183	1.7136	1.7048
MASE	0.7046	0.7305	0.6765	0.6520	0.6737	0.6335	0.6610	0.6087
PB of MASE INARMA-LT/Benchmark	0.4965	0.4841	0.4584	0.3937	0.4533	0.3744		
PB of MASE INARMA-h/Benchmark	0.6060	0.5935	0.5628	0.4786	0.5587	0.4411		
RGRMSE INARMA-LT/Benchmark	33.7864	1.0793	1.1154	1.2958	1.1179	1.3940		
RGRMSE INARMA-h/Benchmark	55.3399	0.8959	0.9387	1.0608	0.9408	1.1276		

Table 9-46 Lead-time forecasts $(l = 6)$ for all-INAR(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.4821	-0.6036	-0.2889	-0.0902	-0.2674	0.0809	0.1601	0.1812
MSE	4.6554	5.7420	4.2290	4.3957	4.1954	4.3210	3.0028	3.1387
MASE	8.6367	8.7812	8.1212	7.4458	8.0658	7.0571	5.9356	6.0057
PB of MASE INARMA-LT/Benchmark	0.6020	0.6131	0.5776	0.5511	0.5746	0.5358		
PB of MASE INARMA-h/Benchmark	0.5950	0.6092	0.5680	0.5457	0.5657	0.5275		
RGRMSE INARMA-LT/Benchmark	1.3096	0.8281	0.9230	0.9579	1.2292	1.0088		
RGRMSE INARMA-h/Benchmark	1.3410	0.8558	0.9518	0.9889	1.2503	1.0608		

Table 9-47 Lead-time forecasts $(l = 6)$ for all-INAR(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.3373	-0.4344	-0.1573	0.0399	-0.1373	0.1980	-0.0422	0.1717
MSE	5.4407	6.8987	5.0715	5.5465	5.0477	5.5698	4.2465	4.3950
MASE	1.1607	1.2270	1.1085	1.0758	1.1035	1.0463	1.0254	0.9807
PB of MASE INARMA-LT/Benchmark	0.5209	0.5093	0.4899	0.4399	0.4857	0.4317		
PB of MASE INARMA-h/Benchmark	0.5806	0.5728	0.5476	0.4944	0.5450	0.4768		
RGRMSE INARMA-LT/Benchmark	1.0349	1.0374	1.1060	1.2321	1.1167	1.3288		
RGRMSE INARMA-h/Benchmark	0.9532	0.9344	1.0082	1.0947	1.0153	1.2132		

For $l = 6$, when an all-INAR(1) approach is used, the INARMA-LT forecasts are better than the best benchmark by 28.4 percent using MSE and 15.9 percent using MASE. The INARMA-h results are worse than those of INARMA-LT, but they still outperform the best benchmark by 25.1 percent using MSE and 14.8 percent using MASE.

The results of Table 9-47 show that when only issue points are considered, the INAR(1)-LT results are better than those of INAR(1)-h in terms of MSE. The INARMA forecasts have also smaller MSE and MASE than the best benchmark method.

Table 9-48 Lead-time forecasts $(l = 3)$ for all-INAR(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.2024	-0.3862	0.4237	1.2251	0.4933	1.7622	-0.0196	-0.0042
MSE	12.6178	17.2396	12.3577	16.1478	12.4000	17.6000	11.9562	12.2358
MASE	1.5991	1.8379	1.5665	1.7416	1.5664	1.7936	1.5594	1.5723
PB of MASE INARMA-LT/Benchmark	0.5286	0.5818	0.5020	0.5714	0.5027	0.5849		
PB of MASE INARMA-h/Benchmark	0.5151	0.5671	0.4927	0.5622	0.4948	0.5806		
RGRMSE INARMA-LT/Benchmark	0.9587	0.8979	1.0093	0.9785	1.0148	0.9763		
RGRMSE INARMA-h/Benchmark	1.0324	0.9716	1.0961	1.0560	1.0996	1.0518		

The results for 3,000 series show that when $l = 3$, all-INAR(1)-LT improves the MSE by 3.2 percent and MASE by 0.5 percent compared to the best benchmark. The

INARMA-h is worse than INARMA-LT using MSE and MASE, but it slightly outperforms the best benchmark by one percent using MSE. However, the MASE of INARMA-h is worse than that of SBJ 0.2 by 0.4 percent.

Table 9-49 Lead-time forecasts $(l = 3)$ for all-INAR(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.2533	-0.4123	0.3668	1.1778	0.4357	1.7078	-0.2270	-0.1060
MSE	13.0190	17.7481	12.6853	16.5018	12.7198	17.9045	12.3828	12.6429
MASE	1.6838	1.9347	1.6453	1.8215	1.6447	1.8703	1.6472	1.6584
PB of MASE INARMA-LT/Benchmark	0.5229	0.5795	0.4919	0.5636	0.4911	0.5733		
PB of MASE INARMA-h/Benchmark	0.5098	0.5659	0.4886	0.5569	0.4905	0.5733		
RGRMSE INARMA-LT/Benchmark	1.0066	0.9629	1.0841	1.0607	1.0866	1.0617		
RGRMSE INARMA-h/Benchmark	1.0574	1.0119	1.1363	1.1058	1.1355	1.1061		

The results of Table 9-49 show that when only issue points are considered, the INAR(1)-LT results are still better than those of INAR(1)-h in terms of both MSE and MASE. The INAR(1)-LT forecasts have also smaller MSE and MASE than the best benchmark method.

When $l = 6$, all-INAR(1)-LT improves the MSE by 2.9 percent and MASE by 0.3 percent compared to the best benchmark. Here, the INARMA-h is not only worse than INARMA-LT, but also than the best benchmark by 3.6 using MSE and 1.8 using MASE.

Table 9-50 Lead-time forecasts $(l = 6)$ for all-INAR(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.6658	-1.1833	0.5971	2.1033	0.7374	3.1988	-0.2114	-0.0701
MSE	35.6315	57.2224	33.3025	48.0324	33.3483	52.8916	32.3335	34.5719
MASE	2.5195	3.0745	2.4368	2.8806	2.4360	3.0082	2.4278	2.4818
PB of MASE INARMA-LT/Benchmark	0.5408	0.6026	0.5060	0.5911	0.5081	0.6097		
PB of MASE INARMA-h/Benchmark	0.5109	0.5783	0.4917	0.5720	0.4911	0.5914		
RGRMSE INARMA-LT/Benchmark	1.0251	0.9393	1.1143	1.0638	1.0950	1.0876		
RGRMSE INARMA-h/Benchmark	1.1465	1.0590	1.2257	1.1703	1.2227	1.2285		

Table 9-51 Lead-time forecasts $(l = 6)$ for all-INAR(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.7413	-1.1618	0.5076	2.0657	2.0657	3.1415	-0.5750	-0.2601
MSE	37.5495	59.7750	34.8230	50.0080	50.0080	54.7176	34.2621	36.3839
MASE	2.7269	3.3355	2.6232	3.0858	3.0858	3.2041	2.6189	2.6774
PB of MASE INARMA-LT/Benchmark	0.5420	0.6085	0.4995	0.5841	0.5005	0.5977		
PB of MASE INARMA-h/Benchmark	0.5138	0.5799	0.4905	0.5692	0.4884	0.5851		
RGRMSE INARMA-LT/Benchmark	0.5977	1.0566	0.9881	1.1776	1.1725	1.1722		
RGRMSE INARMA-h/Benchmark	1.1776	1.0943	1.2778	1.2446	1.2821	1.6062		

Based on the results of Table 9-51, when only issue points are considered, the INAR(1)-LT still produces better forecasts than INAR(1)-h in terms of both MSE and MASE. The INAR(1)-LT forecasts have also smaller MSE and MASE than the best benchmark method.

In general, the results of this section show that an all-INAR(1) method always yields much more accurate lead time forecasts. This is true for both data sets and lengths of lead time. The INAR(1)-LT forecasts are more accurate than the INAR(1)-h forecasts; however, the latter still outperforms the benchmarks in most cases.

9.6 Conclusions

In this chapter, the results of the empirical analysis have been presented. The purpose of the empirical analysis was to assess the empirical validity of the findings suggested by theoretical and simulation results.

Analysing two data sets with different properties (such as length of history and the sparsity of data) enabled us to investigate the sensitivity of the results to such factors. Both data sets have been filtered to eliminate lumpy series.

Four approaches to determine the autoregressive and moving average orders of INARMA models have been tested: treating all series as INAR(1) processes, treating all series as INARMA(1,1) processes, and one-stage and two-stage identification procedures. The results show that although the two-stage identification produces the best results, treating all as INAR(1) also seems a promising approach.

The INARMA one-step ahead forecasts have been compared to the benchmark methods which are: Croston, SBA and SBJ with smoothing parameters 0.2 and 0.5. The results show that for the sparser data set with longer history (called the 16,000 series), there is a substantial improvement in using INARMA over the benchmarks in terms of MSE and MASE. However, for the faster intermittent series with shorter history (called the 3,000 series), the improvement is narrow. The simulation results in the previous chapter also confirm that for short length of history, only when the autoregressive parameter is high, the INARMA method outperforms the benchmarks. However, for low autoregressive parameters, the superiority of INARMA method is subject to availability of more observations.

When the forecasts are made for *h*-step ahead, the improvement of INARMA over benchmarks is greater compared to the one-step ahead forecasts, confirming the simulation results.

The effect of using different estimation methods on the accuracy of INARMA forecasts has also been tested. These methods include CLS, YW and CML for the INAR(1) model, and YW and CLS for INMA(1) and INARMA(1,1) models. The empirical results confirm the simulation results in that for INAR(1) series with low autoregressive parameter (which is true for both 16,000 and 3,000 series), YW estimation method produces better results than CLS and CML. The YW estimates are also better than CLS for an INMA(1) model. But for an INARMA(1,1) model, the CLS estimates produces better forecasts than YW, agreeing with simulation results.

In order to assess the sensitivity of forecasts to the length of history, the 16,000 data set which contains 72 periods of demand data has been used. The results show that when more observations are available, all of the forecasting methods improve. However, the benchmark methods improve at a greater rate than INARMA.

The empirical results of this chapter also show that for very sparse data, the absolute error (AE) of the naïve method is very close to zero which results in very high MASE for all of the competing forecasting methods. When only issue points are considered, because the chance of observing a zero demand after a positive demand is high, the AE of the naïve method increases considerably which results in very small MASE for INARMA and other benchmark forecasts. Therefore, it seems that for highly intermittent data, the MASE does not produce reliable results.

The lead time forecasts are also compared for INARMA and benchmark methods. Two approaches have been followed regarding INARMA lead time forecasts. The first approach is based on the conditional expected value of the lead time aggregated INARMA model. The second approach is based on the cumulative *h*-step ahead forecasts. The results show a considerable improvement over benchmark methods in terms of MSE and MASE for both approaches with the former approach producing better results than the latter.

Chapter 10 CONCLUSIONS AND FURTHER RESEARCH

10.1 Introduction

This chapter summarizes the contributions and conclusions of this PhD thesis. Also, the limitations of this research are identified and future research avenues are suggested.

The initial problem of this research was to assess the potential gain (in forecast accuracy) by modelling and forecasting intermittent demand using INARMA models compared to simpler methods.

In order to solve the above problem, the following research questions were identified in chapter 1:

1. How can the appropriate integer autoregressive moving average (INARMA) model be identified for a time series of counts?
2. How can the parameters of integer autoregressive moving average (INARMA) models be estimated?
3. How can an INARMA process be forecasted over a lead time?
4. Do INARMA models provide more accurate forecasts for intermittent demand than non-optimal smoothing-based methods?

All of the above questions have been answered and the contributions of this thesis are summarized in the following section.

10.2 Contributions

The contributions of this thesis are as follows:

- The unconditional variance and the autocorrelation function of an INARMA(p,q) process are found in this research. It is shown that the autocorrelation and partial autocorrelation functions of an INARMA(p,q) process have the same structure as those of ARMA processes. Therefore, the estimates of these functions can be used to identify the moving average and autoregressive orders of the process, respectively (question 1).
- Two automated identification approaches are also suggested. A two-stage approach first uses a Ljung-Box test to find if the series has any serial dependence. Then the AIC is used to select among the possible INARMA models. On the other hand, a one-stage approach ignores the first step and only uses the AIC. These two approaches are compared in terms of the percentage of series for which the INARMA model is identified correctly and also in terms of the accuracy of forecasts. The results show that although the two stage method performs better when data is i.i.d. Poisson, for other INARMA models the one-stage method produces better results (question 1).
- The effect of misidentification on the forecast accuracy is checked. It is found that misidentification has a great impact if the autoregressive parameter is high. However, for moving average processes, this effect is smaller (question 1).
- It has been found, through simulation and empirical analysis, that ignoring

the identification step and forecasting with an INAR(1) or an INARMA(1,1) method produces promising results. It is shown that the former outperformers the latter both for simulation and empirical data (question 1).

- Three estimation methods are used in this research namely, YW, CLS and CML (CML only for INAR(1) process). The performance of these estimators has been compared in the literature in terms of the bias and MSE of the parameter estimates (Al-Osh and Alzaid, 1987; Bu, 2006; Bu et al., 2008). These studies only consider large sample sizes, which is not common for intermittent demand data. Therefore, we look at the performance of these estimators for small samples including $n = 24, 36$. Our comparison also includes the impact of different estimation methods on forecast accuracy. It is shown that large differences in accuracy of parameter estimates do not necessarily translate to large differences in forecasting accuracy (question 2).
- As shown in chapter 5, finding the ACF of the INARMA(p,q) process has enabled us to obtain the YW estimates for such processes. These estimators are derived for the INARMA(1,1) and INARMA(2,2) processes (question 2).
- It is shown that the aggregation of an INARMA(p,q) process over lead time results in an INARMA(p,q) process with the same autoregressive and moving average parameters but with a different innovation parameter. The conditional expected value of the aggregated process is obtained which is used for lead time forecasting (question 3).
- INARMA models have had applications in many areas including medical science and economics. This research is the first attempt to use these models for intermittent demand modelling and forecasting. In order to find if there is any benefit in terms of forecast accuracy in using INARMA methods, the forecasts are compared to those of some benchmark methods namely, Croston’s method, SBA, and SBJ method. The simulation results show that when data has a high autocorrelation, there is a considerable improvement. The validity of this result is also confirmed by two empirical data sets (question 4).

10.3 Conclusions from the Theoretical Part of the Thesis

In this section the main theoretical findings of this PhD research are discussed.

10.3.1 The Unconditional Variance of an INARMA(*p*,*q*) Model

As discussed in chapter 3, although many studies focused on finding the stochastic properties of INAR(*p*) and INMA(*q*) processes, the same is not true for the mixed INARMA(*p*,*q*) process. It can be easily shown that, assuming stationarity, the unconditional mean of the INARMA(*p*,*q*) process is given by:

$$E(Y_t) = \left(\frac{1+\sum_{j=1}^{q}\beta_j}{1-\sum_{i=1}^{p}\alpha_i}\right)\mu_Z$$

Equation 10-1

where μ_Z is the mean of the innovations in Equation 3-50. As the first result of this research, the unconditional variance of this process is obtained.

$$\begin{aligned}\mathrm{var}(Y_t) = {} & \frac{\mu_Z}{1-\sum_{i=1}^{p}\alpha_i^2}\left[\frac{1+\sum_{i=1}^{q}\beta_i}{1-\sum_{i=1}^{p}\alpha_i}\sum_{i=1}^{p}\alpha_i(1-\alpha_i)+\sum_{i=1}^{q}\beta_i(1-\beta_i)\right] \\ & +\frac{\sigma_Z^2}{1-\sum_{i=1}^{p}\alpha_i^2}\left[1+\sum_{i=1}^{q}\beta_i^2+2\sum_{i=1}^{\min(p,q)}\alpha_i\beta_i\right] \\ & +\frac{2\sum_{j=1}^{p-1}\sum_{i=1}^{p-j}\alpha_i\alpha_{i+j}\gamma_j+2\sum_{i=1}^{\min(p,q)}\sum_{j=i+1}^{q}\alpha_i\beta_j\gamma_{j-i}^{YZ}}{1-\sum_{i=1}^{p}\alpha_i^2}\end{aligned}$$

Equation 10-2

where γ_k can be obtained in terms of $\mathrm{var}(Y_t)$ (or γ_0) from Equation 10-4 and γ_k^{YZ} is the cross-covariance at lag k derived in Appendix 3.C to be:

$$\gamma_k^{YZ} = \beta_k\sigma_Z^2+\sum_{i=1}^{k}\alpha_i\gamma_{k-i}^{YZ}$$

Equation 10-3

It can be seen that when $q=0$, the unconditional mean and variance of the INAR(*p*) process are not equal for the case of Poisson innovations, confirming the literature (Bu and McCabe, 2008). As explained in section 3.3.2, this is because of the extra assumption of independence between thinning operations at different times made by Du and Li (1991).

When $p=0$, the Equation 10-2 reduces to the unconditional variance of an INMA(*q*) process agreeing with the result of Brännäs and Hall (2001). It can be shown that, for Poisson innovations, the mean and variance of the INMA(q) process are equal which again confirms the results in the literature (Brännäs and Hall, 2001).

10.3.2 The Autocorrelation Function of an INARMA(p,q) Model

The autocovariance of an INARMA(p,q) process has been found in section 3.3.8.2 to be:

$$\gamma_k = \begin{cases} \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} + \beta_k\gamma_0^{YZ} + \beta_{k+1}\gamma_1^{YZ} + \cdots + \beta_q\gamma_{q-k}^{YZ} & k \le q \\ \alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} & k > q \end{cases}$$

Equation 10-4

where γ_k^{YZ} is the cross-covariance at lag k given by Equation 10-3. Therefore, as the second finding of this research, the autocorrelation function of an INARMA(p,q) process is as follows:

$$\rho_k = \begin{cases} \dfrac{\alpha_1\gamma_{k-1} + \alpha_2\gamma_{k-2} + \cdots + \alpha_p\gamma_{k-p} + \beta_k\gamma_0^{YZ} + \beta_{k+1}\gamma_1^{YZ} + \cdots + \beta_q\gamma_{q-k}^{YZ}}{\gamma_0} & k \le q \\ \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} + \cdots + \alpha_p\rho_{k-p} & k > q \end{cases}$$

Equation 10-5

where γ_0 is the unconditional variance of the process and is given by Equation 10-2.

It is shown that if $q = 0$, the Equation 10-5 reduces to the ACF of an INAR(p) process given by Du and Li (1991). Also if $p = 0$, the Equation 10-5 provides the ACF of an INMA(q) process which is given by Brännäs and Hall (2001) (see section 3.3.8.2).

It is also shown is section 4.3.1 that the ACF structure of the INARMA(p,q) process is analogous to that of an ARMA(p,q) process. Also, based on the discussion in section 4.3.2.3, the partial autocorrelation function of an INARMA(p,q) process is the analogue of that of an ARMA process, is infinite and behaves like the PACF of a pure integer moving average process.

10.3.3 The YW Estimators of an INARMA(1,1) Model

Based on Equation 10-2 and Equation 10-5, we have found the YW estimators for an INARMA(1,1) process to be:

$$\hat{\alpha} = \frac{r_2}{r_1} = \frac{\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{t=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}$$

Equation 10-6

$$\hat{\beta} = \frac{(1+\hat{\alpha})(\hat{\alpha} - r_1)}{r_1(1+3\hat{\alpha}) - 1 - \hat{\alpha} - 2\hat{\alpha}^2}$$

Equation 10-7

where r_k is the estimate of the autocorrelation at lag k, ρ_k. The innovation parameter is then estimated from the expected value of the process

$$\hat{\lambda} = \frac{1-\hat{\alpha}}{1+\hat{\beta}} \frac{\sum_{t=1}^{n} Y_t}{n}$$

Equation 10-8

We have also derived the YW estimators for an INARMA(2,2) process for presentational purposes (see section 5.8).

10.3.4 Lead Time Forecasting of an INARMA(*p*,*q*) Model

It is shown in chapter 6 that aggregation of an INARMA(p,q) process over a lead time results in an INARMA(p,q) process. The aggregated process has the same autoregressive and moving average parameters but a different innovation parameter. When the innovations of the original process are $Z_t \sim Poi(\lambda)$, the innovations of the aggregated process are $Z_\tau \sim Poi((l+1)\lambda)$.

It is also found that the aggregated process can be written in terms of the last p observations as follows:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Y_{t-1} + \cdots + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^p} \psi_{ij}^p \circ Y_{t-p+1} + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^{p+1}} \psi_{ij}^{p+1} \circ Z_{t+k_{ij}}$$

Equation 10-9

where all the parameters are defined in Table 6-4. The above result is then used to find the conditional expected value of the aggregated process which is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_{t-p+1},\dots,Y_{t-1},Y_t\right) = \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2\right) Y_{t-1} + \cdots + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^p} \psi_{ij}^p\right) Y_{t-p+1} + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^{p+1}} \psi_{ij}^{p+1}\right)\lambda$$

Equation 10-10

Based on the above results, the lead time forecasts for the specific INARMA processes used in this research are derived as follows:

$$E\left[\sum_{i=1}^{l+1} Y_{t+i} \mid Y_t\right] = \begin{cases} (l+1)\lambda & \text{for INARMA}(0,0) \\ \dfrac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \dfrac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right] & \text{for INAR}(1) \\ (l+1)(1+\beta)\lambda & \text{for INMA}(1) \\ \dfrac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \dfrac{\lambda(1+\beta)}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1}\alpha^j\right] & \text{for INARMA}(1,1) \end{cases}$$

Equation 10-11

10.4 Conclusions from the Simulation Part of the Thesis

In this section, the main findings of the simulation part of this research are summarized. Four INARMA models are used in this thesis which are: INARMA(0,0), INAR(1), INMA(1) and INARMA(1,1). The range of parameters chosen for these processes is shown in Table 7-1.

10.4.1 The Performance of Different Estimation Methods

The performance of YW, CLS, and CML (only for INAR(1) process) is compared. The simulation results show that, for an INAR(1) process, the MSE of estimates produced by CML is generally less than that of YW and CLS. However, the accuracy of forecasts produce by CML is not very far from those by YW and CLS. The YW and CLS estimators are close when the number of observations is high. But, for small samples, the difference is high when the autoregressive parameter is high. In such cases, CLS produces much better estimates for both α and λ in terms of MSE than YW. On the other hand, for small values of α, YW results in better estimates. This is also true for the accuracy of forecasts produced by these estimation methods.

For an INMA(1) process, for small number of observations, CLS generally has smaller MSEs than YW except for very high values of β. For large number of observations, YW has smaller MSEs than CLS for high values of β. However, this does not have a great effect on the accuracy of forecasts produced by each method. The MSE of $\hat{\lambda}$ for both YW and CLS estimates increases with an increase in β but the same is not necessarily true for the MSE of $\hat{\beta}$.

Finally, for an INARMA(1,1) process, CLS generally produces better estimates especially when the number of observations is small and the autoregressive parameter is high. This is also true for the accuracy of forecasts produced by CLS compared to those by YW.

The results of three-step and six-step ahead forecasts show that the forecasts based on YW and CLS estimation methods are generally very close for all of the three INARMA processes, but YW-based results are slightly better in many cases.

10.4.2 The Croston-SBA Categorization

Syntetos et al. (2005) provide a categorization scheme for Croston and SBA based on MSE to establish the areas that each method should be used over the other. They use the squared coefficient of variation (CV^2) of demand size and the average inter-demand interval (p) to identify different regions shown in Figure 8-1. Because this categorization is based on the assumption that demand occurs as an i.i.d. Bernoulli process, it is worth testing if it also holds for an i.i.d. Poisson (an INARMA(0,0)) process. The simulation results show that not only the Croston-SBA categorization based on MSE holds for i.i.d. Poisson demand, but it also generally holds using MASE.

The simulation results also show that when the number of observation increases, the advantage of SBA over Croston decreases. This is also shown by direct mathematical calculations in section 8.5.

$$\mathrm{MSE}_{\mathrm{SBA}} - \mathrm{MSE}_{\mathrm{Croston}} \approx \left[\left(1-\frac{\alpha}{2}\right)^2 - 1\right]\left[\frac{\alpha + 2(1-\alpha)^{2n+1}}{2-\alpha}\right]\left[\frac{(p-1)^2\mu^2}{p^4} + \frac{\sigma^2}{p^2}\right]$$

Equation 10-12

Based on the Equation 10-12, when the smoothing parameter is small ($\alpha = 0.2$), the above coefficient decreases when n increases; therefore the difference between MSE of Croston and SBA also decreases. However, because the above coefficient reaches a limit of $\left(\frac{\alpha^2}{4} - \alpha\right)\left(\frac{\alpha}{2-\alpha}\right)$, the advantage of SBA over Croston does not change noticeably when the number of observations is high. On the other hand, when the smoothing parameter is high ($\alpha = 0.5$), the difference between the MSE of Croston and SBA changes little with changes in n.

Although the Croston-SBA categorization is for i.i.d. demand, we have also tested it when demand is produced by an INAR(1), INMA(1) or an INARMA(1,1) process. The results show that the Croston-SBA categorization generally holds for all of these processes.

10.4.3 Identification in INARMA Models

We test the two identification approaches mentioned in section 10.2. The simulation results show that when the data is produced by an INARMA(0,0) process, the two-stage method produces better results than the one-stage method. This is true in terms of the percentage of time that the model is identified correctly and also in terms of the accuracy of forecasts. However, when data is produced by INAR(1), INMA(1) or

INARMA(1,1) processes, the one-stage method outperforms the two-stage method.

For INAR(1) and INARMA(1,1) processes, when the autoregressive parameter is low, the process is misidentified in most cases for both identification methods. But when the autoregressive parameter is high, the performance of both identification methods improves. Obviously, when more observations are available, the percentage of correct identification increases for both methods.

The INMA(1) process is misidentified in most of the cases. However, the results show that it does not affect the forecasting accuracy to a great extent. In general, the performance of both identification methods improves for higher moving average parameters and longer length of history.

The simulation results also show that, for INAR(1) and INARMA(1,1) processes, misidentification has a great impact on the accuracy of forecasts when the autoregressive parameter is high. When the autoregressive parameter is small, or the process is an INMA(1) process, the effect of misidentification on forecasting accuracy is small.

We also test the case that the identification step is ignored and an INAR(1) or an INARMA(1,1) method is used for forecasting. The results show that when the number of observations is small, this approach produces better forecasts for INAR(1) and INMA(1) series. The results also show that the all-INAR(1) approach produces better forecasts than the all-INARMA(1,1) approach for all four INARMA series.

10.4.4 Comparing INARMA with the Benchmark Methods

The results show that when the order of the INARMA model is known, the INARMA method almost always produces the lowest MSE when all points in time is considered. When only issue points are considered, the INARMA forecasts are biased and therefore are not always better than the benchmark methods.

For INARMA(0,0) and INMA(1) processes, the improvement over benchmarks is not considerable. However, for the INAR(1) and INARMA(1,1) processes, the improvement is considerable when the autoregressive parameter is high. This is true for both MSE and MASE of forecasts.

The degree of improvement (by using INARMA over the benchmark methods) for the h-step ahead forecasts does not change for INARMA(0,0) process. But for an INMA(1) process, the performance of INARMA compared to benchmark methods is improved for h-step ahead forecasts compared to one-step ahead forecasts. For INAR(1) and INARMA(1,1) processes, the performance of INARMA over the benchmark methods is improved compared to the one-step ahead case when the autoregressive parameter is low. But when the autoregressive parameter is high, the fact that the forecasts converge to the mean of the process results in poor forecasts compared to the one-step ahead case.

All of the above results were for the case that the order of the INARMA model is known. This is obviously not true in practice and the INARMA model needs to be identified. The results of identification show that treating all as an INAR(1) model is a promising approach especially for high autoregressive parameters and short length of history. Therefore, we compare an all-INAR(1) method with the benchmark methods. The results show that for INARMA(0,0) and INMA(1) processes, the benchmark methods outperform INARMA especially for more sparse demand. For INARMA(1,1) series, when the autoregressive parameter is high, INARMA is considerably better than the benchmark methods in terms of MSE and MASE. Obviously, the results for INAR(1) series are the same as the case that the order is known.

We also compare the lead time forecasts produced by an all-INAR(1) method with the benchmark methods. The results show that, for INARMA(0,0) and INMA(1) series, the all-INAR(1) forecasts are better than the best benchmark in most of the cases. For INAR(1) and INARMA(1,1) series, when the autoregressive parameter is high, the improvement of INARMA over the best benchmark is narrow for short length of history. However with more observations, the improvement also increases. For small autoregressive parameters, the INAR(1) method always outperforms the benchmark methods. Again, the improvement increases with an increase in the length of history.

10.5 Conclusions from the Empirical Part of the Thesis

The simulation analysis of this research compared the accuracy of forecasts (in terms of MSE and MASE) by an INARMA method with those produced by benchmark methods when the data is generated by an INARMA model. Empirical analysis is performed to validate the theoretical and simulation results for real data. Two data sets used in this research are the Royal Air Force (RAF) individual demand histories of 16,000 SKUs over a period of 6 years (monthly observations) and demand history of 3,000 SKUs from the automotive industry over a period of 2 years (24 months). Both data sets are filtered to eliminate lumpy series, since the INARMA models with Poisson marginal distribution are not appropriate for lumpy demand. The PB of MASE and the RGRMSE are used to measure the forecast error in addition to ME, MSE and MASE.

10.5.1 Identification in INARMA Models

The two identification methods (two-stage and one-stage) are compared. The empirical results for both 16,000 and 3,000 series show that the two-stage method produces more accurate forecasts. This is not surprising because the identification shows that the majority of series in both data sets are identified as INARMA(0,0) and the simulation results show that for INARMA(0,0) series the two-stage method is better than the one-stage method.

The empirical results also confirm the simulation results that ignoring the identification step, and forecasting with an INARMA(1,1) or an INAR(1) process, is a promising approach. It is also confirmed that the latter outperforms the former.

10.5.2 The Performance of Different Estimation Methods

After the appropriate INARMA model is identified for all the series in each data set, different INARMA series are separated and forecasted by the corresponding INARMA method. The performance of different estimation methods in terms of the accuracy of forecasts are tested for each INARMA model.

The empirical results agree with the simulation results for all three INARMA models. For INAR(1) series (with α around 0.1 and λ around 2), the YW produces better forecasts than CLS and CML agreeing with the simulation results for similar parameters. For INMA(1) series, YW produces slightly better results than CLS. However, for INARMA(1,1) series, CLS outperforms YW, which again confirms the simulation results.

10.5.3 Comparing INARMA with the Benchmark Methods

The forecast accuracy of INARMA (based on all identification methods) is compared to that of the benchmark methods for both 16,000 and 3,000 series. The results show that, for the former data set, there is an improvement by using INARMA methods for one-step ahead forecasts for all accuracy measures. The improvement is narrow for 3,000 series which is expected because of the short length of history and low autocorrelation. This slight improvement is based on MSE and, in fact, the MASE of INARMA is worse than the benchmarks for 3,000 series.

The accuracy of h-step ahead INARMA forecasts are even better than the benchmarks compared to the one-step ahead forecasts. This is true for both data sets.

The lead time forecasts are obtained based on the conditional expected value of the aggregated INARMA process and also from the cumulative h-step ahead forecasts. The empirical results show that for INAR(1) series of both data sets, both of these methods produce considerably better forecasts than the benchmark methods with the former outperforming the latter. The same is true for INMA(1) series of both data sets but not for INARMA(1,1) series. For INARMA(1,1) series, the cumulative h-step ahead forecasts are better than those based on conditional expected value of the aggregated process. It should be borne in mind that the number of INARMA(1,1) for both data sets are very small.

10.5.4 The Problem with MASE

The empirical results show that when the data series is highly intermittent (which is the case for most of the 16,000 series), the MASE of all forecasting methods is very high. This would suggest that all methods are worse than naïve. Because the data series contain many zeros in the estimation period, the error of naïve in most of the time periods is zero. Therefore the in-sample MAE is very small and the MASE is very large. This is true when all points in time are considered.

However, for the case of issue points, because it is more likely that a nonzero demand is followed with a zero demand, the absolute error of naïve and therefore the in-sample MAE is large. As a result, the MASE of the forecasting methods is smaller compared to the all points in time case. The empirical results confirm this. As a result, for highly intermittent data, MASE does not provide reliable results.

10.6 Practical and Software implications

In this thesis we have applied an INARMA method to model and forecast non-erratic intermittent demand (see the classification by Boylan et al., 2008). Four models have been assumed to include autoregressive, moving average, mixed models and also an i.i.d. Poisson process. All simulation and empirical results are based on the following assumptions:

- The intermittent demand can be modelled with either an INARMA(0,0), INAR(1), INMA(1) or an INARMA(1,1) process with Poisson marginal
- The benchmark methods to compete against INARMA method are: Croston, SBA and SBJ
- The forecast accuracy measures are: ME, MSE, MASE, RGRMSE, and PB, with the last two only used for empirical analysis

It has been shown that INARMA performs best when data has high autocorrelation. Even with low autocorrelation INARMA outperforms the benchmarks when the length of history is large. However, we cannot claim that when the above assumptions are violated, these results still hold. For example, the empirical data of this research were filtered to comply with the first assumption.

Therefore, in order to use the INARMA method, the distribution that fits the demand data of the organization should first be found. All discrete self-decomposable distributions can be used as marginal distributions of INARMA models. This includes: Poisson, generalized Poisson, and negative binomial distributions. The generalized Poisson allows for both under and overdispersion (Brännäs, 1994), and the negative binomial allows for overdispersion (McCabe and Martin, 2005).

The results of this research suggest that although identification of the autoregressive and moving average orders of the INARMA models results in more accurate forecasts, using simple models such as an INAR(1) model also produces good forecasts. This is an especially useful method when the length of history is not long enough for identification. However, it should be mentioned that the simulation results of this research were based on limited INARMA models and the empirical data did not support higher order models. Therefore, for higher order models, the performance of using a simple model instead of identification needs to be tested first through simulation.

The YW and CLS estimation methods are not based on distributional assumptions and therefore can be used for all cases. However, the maximum likelihood estimators for the corresponding distribution should be obtained. The results of this research can be used, at least for the specific INARMA models and parameters, as a guide to when one estimation method should be used over another.

Although the lead time forecasts in this thesis are restricted to Poisson, the findings can be easily amended for other distributions.

The conventional ARIMA models are offered by standard forecasting programmes such as Autobox (Automatic Forecasting Systems) and Forecast Pro (Business Forecast Systems). These programmes provide automated identification, estimation and forecasting for ARIMA models. For instance, Forecast Pro uses the Bayesian information criterion (BIC) along with some other rules for identification and unconditional least squares for estimation (Goodrich, 2000). Autobox matches the SACF with theoretical ACF for some starting ARIMA models and then uses AIC to select the best model (Reilly, 2000).

Although INARMA models have had applications in many areas (see section 3.3.9), there is currently no software based on these models. This research suggested a new application area for INARMA models in intermittent demand forecasting. The improvement in forecast accuracy by using INARMA over the benchmark methods such as Croston (which is offered in some forecasting programmes such as Forecast Pro), also urges the need to develop forecasting software for these models.

In doing so, this research can be considered a starting point. We have addressed issues of whether AIC produces satisfactory results, which estimation method is best in what case, and h-step ahead and lead time forecasting for a limited number of INARMA models.

10.7 Limitations and Further Research

Throughout this research we have made assumptions that can be relaxed in future studies. Although most of the theoretical findings are not based on any marginal distribution, for simulation and empirical analyses we restricted the research to a specific distribution. As previously discussed in section 7.2, the Poisson distribution was selected due to its interesting properties. However, it does not allow for overdispersion in data and in such cases other distributions should be assumed. For example, Alzaid and Al-Osh (1993) show that the Generalized Poisson INAR(1) process has more variability than PoINAR(1) because its variance is twice that of PoINAR(1). This could be useful for modelling moderately lumpy demand.

Other adaptations of INARMA models to take into account non-stationarity, seasonality and trend can also be used. For example the signed binomial thinning models introduced by Kim and Park (2008) allow for negative values and negative autocorrelations and also use the differencing operator to remove the trend and seasonality.

As mentioned in section 7.3.1, the simulation and empirical results of this research have been based on INARMA models with $p, q \leq 1$. A natural extension would be to use higher order models. As seen in chapter 9, our empirical data did not support such models; however, other data sets may be better fitted by higher order INARMA models.

This research has only focused on aggregation of INARMA models over lead time. The temporal and cross-sectional aggregation in INARMA models were briefly reviewed in chapter 3. A problem is identified with the result of temporal aggregation of an INAR(1) process provided by Brännäs et al. (2002). More studies should be carried out in both temporal and cross-sectional aggregation fields.

As mentioned in section 4.5, the complexity of the likelihood function has restricted the use of penalty functions for INARMA models. We have used the AIC based on the likelihood function of ARIMA models and shown that the results are satisfactory. However, there is still a need for finding the AIC based on INARMA models.

As confirmed in chapters 8 and 9, the INARMA forecasts are biased when only issue points are considered. The estimates, and as a result, forecasts, can be revised to reduce such bias. For example, the CLS of an INAR(1) process is based on the following criterion:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [Y_t - E(Y_t|Y_{t-1})]^2$$

Equation 10-13

where $\boldsymbol{\theta} = (\alpha, \lambda)'$. Now, when only issue points are considered, the estimates of the parameters will be updated after observing a positive demand. Therefore, in Equation 10-13 the last observation which is Y_n is definitely positive and its conditional expected value should be $E(Y_n|Y_{n-1}, Y_n > 0)$ instead of $E(Y_n|Y_{n-1})$. As a result, the new least squares criterion will be:

$$Q_n'(\boldsymbol{\theta}) = \sum_{t=1}^{n-1} [Y_t - E(Y_t|Y_{t-1})]^2 + [Y_n - E(Y_n|Y_{n-1}, Y_n > 0)]^2$$

Equation 10-14

The revised CLS estimates of parameters can be obtained by minimization of the above criterion. Numerical methods are needed to find these estimates. The new estimates can be compared to CLS estimates in terms of their impact on forecast accuracy when only issue points are considered. Other estimation methods that allow for such revisions such as maximum likelihood could also be considered.

Another limitation of this study is that we have used the conditional expectation to produce forecasts. This enabled us to compare the MMSE INARMA forecasts to point forecasts from benchmark methods. The natural next step is to forecast the whole distribution instead. This has been done in a number of studies in the INARMA literature (Freeland and McCabe, 2004b; McCabe and Martin, 2005; Bu and McCabe, 2008). This will then enable us to compare INARMA method with bootstrapping methods in an IDF context.

Incorporating explanatory variables in INARMA models has been the subject of some studies (Brännäs, 1995; Brännäs and Quoreshi, 2004). As discussed in section 2.3.3, causal models for IDF have not yet been well developed in the literature and the integration of these models with INARMA models would be an interesting line of research.

The benchmark methods of this research were Croston, SBA and SBJ. We have used two arbitrary values for the smoothing parameter of these methods. This could be replaced by the optimum smoothing parameter. Also, different smoothing parameters could be used for updating demand size and inter-arrival time.

Although this study has only focused on forecasting and not inventory control, it is expected that the improvement in the mean (forecasts) would translate to better percentile estimates and better inventory results. Obviously, this could be tested in future studies to find whether using INARMA method would result in better inventory measures such as service level and inventory level that have been suggested by Teunter and Duncan (2009).

Different inventory models need different estimation of parameters, i.e. mean, percentiles, or estimates of demand sizes and inter-demand invervals. Syntetos et al. (2008) develop a modified periodic order-up-to-level inventory policy for intermittent demand which relies upon demand sizes and inter-demand intervals. Depending on what inventory model is applied, INARMA methods can provide all of these estimates. For example, in the case where demand follows an INAR(1) model:

$$E(\text{demand size}) = \frac{\lambda/(1-\alpha)}{1-e^{-\lambda/(1-\alpha)}}$$

Equation 10-15

$$E(\text{demand interval}) = \frac{1-\alpha}{\lambda}$$

Equation 10-16

Finally, the empirical results of this research have been restricted to two data sets. More empirical analyses can be done to further assess the sensitivity of results to the length of history and the level of intermittence.

References

Adya, M., F. Collopy, J. S. Armstrong and M. Kennedy (2001). Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting* **17**(2): 143-157.

Aggoun, L., L. Benkherouf and L. Tadj (1997). A hidden Markov model for an inventory system with perishable items. *Journal of Applied Mathematics and Stochastic Analysis* **10**(4): 423-430.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. B. N. Petrov and F. Csaki (Eds.), Proceedings of the second International Symposium on Information Theory, Akadémiai Kaidó, Budapest.

Al-Osh, M. A. and A. A. Alzaid (1987). First order integer-valued autoregressive (INAR(1)) processes. *Journal of Time Series Analysis* **8**(3): 261-275.

Al-Osh, M. A. and A. A. Alzaid (1988). Integer-valued moving average (INMA) process. *Statistical Papers* **29**(4): 281-300.

Aly, E.-E. A. A. and N. Bouzar (1994). On some integer-valued autoregressive moving average models. *Journal of Multivariate Analysis* **50**(1): 132-151.

Aly, E.-E. A. A. and N. Bouzar (2005). Stationary solutions for integer-valued autoregressive processes. *International Journal of Mathematics and Mathematical Sciences* **2005**(1): 1-18.

Alzaid, A. A. and M. Al-Osh (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Statistica Neerlandica* **42**(1): 53-61.

Alzaid, A. A. and M. Al-Osh (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability* **27**(2): 314-324.

Alzaid, A. A. and M. Al-Osh (1993). Some autoregressive moving average processes with generalized Poisson marginal distributions. *Annals of the Institute of Statistical Mathematics* **45**(2): 223-232.

Amemiya, T. and R. Y. Wu (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* **67**(339): 628-632.

Berglund, E. and K. Brännäs (2001). Plants' entry and exit in Swedish municipalities. *The Annals of Regional Science* **35**(3): 431-448.

Blundell, R., R. Griffith and F. Windmeijer (2002). Individual effects and dynamics in count data models. *Journal of Econometrics* **108**(1): 113-131.

Böckenholt, U. (1999). Mixed INAR(1) Poisson regression models: Analyzing heterogeneity and serial dependencies in longitudinal count data. *Journal of Econometrics* **89**(1-2): 317-338.

Bookbinder, J. H. and A. E. Lordahl (1989). Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions* **21**(4): 302-312.

Bowerman, B. L., R. T. O'Connell and A. B. Koehler (2005). *Forecasting, time series, and regression*, 4th ed., Thomson Brooks/Cole: Belmont, CA.

Box, G. E. P. and G. M. Jenkins (1970). *Time series analysis: Forecasting and control*, Holden-Day: San Francisco.

Box, G. E. P., G. M. Jenkins and G. C. Reinsel (1994). *Time series analysis: Forecasting and control*, 3rd ed., Prentice Hall: New Jersey.

Boylan, J. E. and A. A. Syntetos (2006). Accuracy and accuracy implication metrics for intermittent demand. *FORESIGHT: International Journal of Applied Forecasting* (4): 39-42.

Boylan, J. E. and A. A. Syntetos (2007). The accuracy of a modified Croston procedure. *International Journal of Production Economics* **107**(2): 511-517.

Boylan, J. E. and A. A. Syntetos (2008). Forecasting for inventory management of service parts In Complex system maintenance handbook. K. A. H. Kobbacy and D. N. P. Murthy (eds.), Springer: London, 479-508.

Boylan, J. E., A. A. Syntetos and G. C. Karakostas (2008). Classification for forecasting and stock control: a case study. *Journal of the Operational Research Society* **59**(4): 473-481.

Brandt, P. T., J. T. Williams, B. O. Fordham and B. Pollins (2000). Dynamic modeling for persistent event-count time series. *American Journal of Political Science* **44**(4): 823-843.

Brännäs, K. (1994). *Estimation and testing in integer-valued AR(1) models*. Umeå Economic Studies 335, Umeå University.

Brännäs, K. (1995). *Explanatory variables in the AR(1) count data model*. Umeå Economic Studies 381, Umeå University.

Brännäs, K. and A. Hall (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry* **17**(3): 277-291.

Brännäs, K. and J. Hellström (2001). Generalized integer-valued autoregression. *Econometric Reviews* **20**(4): 425-443.

Brännäs, K., J. Hellström and J. Nordström (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting* **18**(1): 19-30.

Brännäs, K. and P. Johansson (1994). Time series count data regression. *Communications in Statistics Theory and Methods* **23**(10): 2907-2925.

Brännäs, K. and S. Quoreshi (2004). *Integer-valued moving average modelling of the number of transactions in stocks*. Umeå Economic Studies 637, Umeå University.

Brewer, K. R. W. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics* **1**(2): 133-154.

Brockwell, P. J. and R. A. Davis (1996). *Introduction to time series and forecasting*, Springer-Verlag: New York.

Brown, R. G. (1959). *Statistical forecasting for inventory control*, McGraw-Hill: New York.

Bu, R. (2006). *Essays in financial econometrics and time series analysis.* PhD thesis: The University of Liverpool, UK.

Bu, R. and B. McCabe (2008). Model selection, estimation and forecasting in INAR(p) models: A likelihood-based Markov Chain approach. *International Journal of Forecasting* **24**(1): 151-162.

Bu, R., B. McCabe and K. Hadri (2008). Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis* **29**(6): 973-994.

Buishand, T. A. (1978). *The binary DARMA(1,1) process as a model for wet-dry sequences*. Technical note 78-01, Department of Mathematics, Agricultural University Wageningen.

Cardinal, M., R. Roy and J. Lambert (1999). On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine* **18**(15): 2025-2039.

Chang, T. J., M. L. Kavvas and J. W. Delleur (1984). Daily precipitation modeling by discrete autoregressive moving average processes. *Water Resources Research* **20**(5): 565-580.

Chatfield, C. and D. L. Prothero (1973). Box-Jenkins seasonal forecasting: Problems in a case-study. *Journal of the Royal Statistical Society, Series A* **136**(3): 295-336.

Collopy, F. and J. S. Armstrong (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science* **38**(10): 1394-1414.

Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society. Series B* **23**(2): 414-422.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly* **23**(3): 289–303.

Drost, F. C., R. v. d. Akker and B. J. M. Werker (2008). Local Asymptotic Normality and efficient estimation for INAR(p) models *Journal of Time Series Analysis* **29**(5): 783-801.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica* **61**(4): 909-927.

Du, J. G. and Y. Li (1991). The integer-valued autoregressive (INAR(p)) model. *Journal of Time Series Analysis* **12**(2): 129-142.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society, Series B* **62**(1): 3-56.

Eaves, A. H. C. and B. G. Kingsman (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society* **55**(4): 431-437.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**(1): 1-26.

Enciso-Mora, V., P. Neal and T. S. Rao (2008). Efficient order selection algorithms for integer-valued ARMA processes. *Journal of Time Series Analysis* **30**(1): 1-18.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting* **8**(1): 81-98.

Franke, J. and T. Seligmann (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their applications to modelling epileptic seizure counts. In <u>Developments in Time Series Analysis</u>. T. S. Rao (Ed.), Chapman and Hall: London, 310-330.

Franke, J. and T. Subba Rao (1995). *Multivariate first order integer valued autoregressions*. Berichte der Arbeitsgruppe Technomathematik, Bericht Nr.95, Universität Kaiserslautern.

Freeland, K. (1998). *Statistical analysis of discrete time series with application to the analysis of workers' compensation claims data.* PhD thesis: University of British Columbia.

Freeland, R. K. and B. P. M. McCabe (2004a). Analysis of low count time series data by Poisson autoregression. *Journal of Time Series Analysis* **25**(5): 701-722.

Freeland, R. K. and B. P. M. McCabe (2004b). Forecasting discrete valued low count time series. *International Journal of Forecasting* **20**(3): 427-434.

Freeland, R. K. and B. P. M. McCabe (2005). Asymptotic properties of CLS estimators in the Poisson AR(1) model. *Statistics and Probability letters* **73**(2): 147-153.

Gardner, E. S. and A. B. Koehler (2005). Comments on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting* **21**(3): 617-618.

Ghobbar, A. A. and C. H. Friend (2002). Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management* **8**(4): 221-231.

Ghobbar, A. A. and C. H. Friend (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers and Operations Research* **30**(14): 2097-2114.

Goodrich, R. L. (2000). The Forecast Pro methodology. *International Journal of Forecasting* **16**(4): 533-535

Goodwin, P. and R. Lawton (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting* **15**(4): 405-408.

Gooijer, J. G. and R. J. Hyndman (2006). 25 years of time series forecasting. *International Journal of Forecasting* **22**(3): 443-473.

Gourieroux, C. and J. Jasiak (2004). Heterogeneous INAR(1) model with application to car insurance. *Insurance: Mathematics and Economics* **34**(2): 177-192.

Graves, S. C. (1999). A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* **1**(1): 50-61.

Grunwald, G. K., R. J. Hyndman, L. Tedesco and R. L. Tweedie (2000). Non-Gaussian conditional linear AR(1) models. *The Australian & New Zealand Journal of Statistics* **42**(4): 479-495.

Hall, B. H., Z. Griliches and J. A. Hausman (1986). Patents and R and D: is there a lag? *International Economic Review* **27**(2): 265-283.

Hamilton, J. D. (1994). *Time series analysis*, Princeton University Press: Princeton, New Jersey.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**(4): 1029-1054.

Harvey, A. C. and C. Fernandes (1989). Time series models for count data or qualitative observations. *Journal of Business & Economic Statistics* **7**(4): 407-417.

Harvey, A. C. and R. G. Pierse (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association* **79**(385): 125-131.

Hua, Z. S., B. Zhang, J. Yang and D. S. Tan (2007). A new approach of forecasting intermittent demand for spare parts inventories in the process industries. *Journal of the Operational Research Society* **58**(1): 52-61.

Hurvich, C. M. and C. L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* **76**(2): 297-307.

Hurvich, C. M. and C. L. Tsai (1995). Model selection for extended quasi-likelihood models in small samples. *Biometrics* **51**(3): 1077-1084.

Hyndman, R. J. (2006). Another look at forecast accuracy metrics for intermittent demand. *FORESIGHT: International Journal of Applied Forecasting* (4): 43-46.

Hyndman, R. J. and A. B. Koehler (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**(4): 679-688.

Jacobs, P. A. and P. A. W. Lewis (1978a). Discrete time series generated by mixtures I: Correlational and runs properties. *Journal of the Royal Statistical Society. Series B* **40**(1): 94-105.

Jacobs, P. A. and P. A. W. Lewis (1978b). Discrete time series generated by mixtures II: Asymptotic properties. *Journal of the Royal Statistical Society. Series B* **40**(2): 222-228.

Jacobs, P. A. and P. A. W. Lewis (1983). Stationary discrete autoregressive moving average time series generated by mixtures. *Journal of Time Series Analysis* **4**(1): 19-36.

Johnston, F. R. and J. E. Boylan (1996). Forecasting for items with intermittent demand. *Journal of the Operational Research Society* **47**(1): 113–121.

Johnston, F. R., J. E. Boylan and E. A. Shale (2003). An examination of the size of orders from customers, their characterisation and the implications for the inventory control of slow moving items. *Journal of the Operational Research Society* **54**(5): 833-837.

Jung, R. C. and A. R. Tremayne (2003). Testing for serial dependence in time series models of counts. *Journal of Time Series Analysis* **24**(1): 65-84.

Jung, R. C. and A. R. Tremayne (2006a). Binomial thinning models for integer time series. *Statistical Modelling* **6**(2): 81-96.

Jung, R. C. and A. R. Tremayne (2006b). Coherent forecasting in integer time series models. *International Journal of Forecasting* **22**(2): 223-238.

Kang, I.-B. (2003). Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data. *International Journal of Forecasting* **19**(3): 387-400.

Karlis, D. (2002). Overdispersed integer valued autoregressive models. COMPSTAT 2002, Berlin, Germany.

Kim, H.-Y. and Y. Park (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers* **49**(3): 485-502.

Kleijnen, J. P. C. and W. Groenendaal (1992). *Simulation: a statistical perspective*, John Wiley & Sons, Inc.: New York, NY, USA.

Klimko, L. A. and P. I. Nelson (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics* **6**(3): 629-642.

Kolassa, S. and W. Schütz (2007). Advantages of the MAD/Mean ratio over the MAPE. *FORESIGHT: The International Journal of Applied Forecasting* (6): 40-43.

Kostenko, A. and R. Hyndman (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society* **57**(10): 1256-1257.

Ksenija, C. (2006). Statistical analysis of wet and dry spells in Croatia by the binary DARMA(1,1) model. *Croatian Meteorological Journal* **41**: 43-51.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics* **22**(1): 79-86.

Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis* **19**(4): 439-455.

Lawrence, A., A. Klimko and P. I. Nelson (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics* **6**(3): 629-642.

Lee, H. L., K. C. So and C. S. Tang (2000). The value of information sharing in a two-level supply chain. *Management Science* **46**(5): 626-643.

Levén, E. and A. Segerstedt (2004). Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics* **90**(3): 361-367.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika* **65**(2): 297-303.

MacDonald, I. L. and W. Zucchini (1997). *Hidden Markov and other models for discrete-valued time series*, Chapman and Hall: London.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting* **1**(2): 111-153.

Makridakis, S. and M. Hibon (1995). *Evaluating accuracy (or error) measures*. Working Paper 95/18/TM, INSEAD

Makridakis, S. and M. Hibon (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting* **16**(4): 451-476.

Man, K. S. (2004). Linear prediction of temporal aggregates under model misspecification. *International Journal of Forecasting* **20**(4): 659-670.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics* **17**(1): 129-136.

McCabe, B. P. M. and G. M. Martin (2005). Bayesian predictions of low count time series. *International Journal of Forecasting* **21**(2): 315-330.

McKenzie, E. (1985). Some simple models for discrete variate series. *Water Resources Bulletin* **21**(4): 645-650.

McKenzie, E. (1986). Autoregressive moving-average processes with negative binomial and geometric marginal distributions. *Advances in Applied Probability* **18**(3): 679-705.

McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability* **20**(4): 822-835.

McKenzie, E. (2003). Discrete variate time series. In Stochastic Processes: Modelling and Simulation. D. N. Shanbhag and C. R. Rao (Eds.), Elsevier: Amsterdam, 573-606.

Mills, T. M. and E. Seneta (1989). Goodness-of-fit for a branching process with immigration using sample partial autocorrelations. *Stochastic Processes and their Applications* **33**(1): 151-161.

Neal, P. and T. S. Rao (2007). MCMC for integer-valued ARMA processes. *Journal of Time Series Analysis* **28**(1): 92-110.

Newbold, P. and C. W. J. Granger (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A* **137**(2): 131-165.

Nijman, T. E. and F. C. Palm (1990). Predictive accuracy gain from disaggregate sampling in ARIMA models. *Journal of Business & Economic Statistics* **8**(4): 405-415.

Ozaki, T. (1977). On the order determination of ARIMA models. *Applied Statistics* **26**(3): 290-301.

Park, Y. and C. W. Oh (1997). Some asymptotic properties in INAR(1) processes with Poisson marginals. *Statistical Papers* **38**(3): 287-302.

Patton, J. D. and R. J. Steele (2003). *Service Parts Handbook,* 2nd Edition, Solomon Pr: Rego Park, NY.

Porras, E. and R. Dekker (2007). An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *European Journal of Operational Research* **184**(1): 101-132.

Porras Musalem, E. (2005). *Inventory theory in practice: Joint replenishments and spare parts control*. PhD thesis: Erasmus University Rotterdam (Tinbergen Institute).

Quddus, M. A. (2008). Time series count data models: An empirical application to traffic accidents. *Accident Analysis & Prevention* **40**(5): 1732-1741

Quenouille, M. H. (1958). Discrete autoregressive schemes with varying time intervals. *Metrika* **1**(1): 21-27.

Rao, A. V. (1973). A comment on "Forecasting and stock control for intermittent demands". *Operational Research Quarterly* **24**(4): 639-640.

Reilly, D. (2000). The AUTOBOX system. *International Journal of Forecasting* **16**(4): 531-533

Rengifo, E. W. (2005). *Econometric analysis of financial count data and portfolio choice: A dynamic approach.* PhD: Université Catholique de Louvain.

Rudholm, N. (2001). Entry and the number of firms in the Swedish pharmaceuticals market. *Review of Industrial Organization* **19**(3): 351-364.

Salas, J. D., C.-h. Chung and B. Fernandez (2001). Relating autocorrelations and crossing rates of continuous and discrete valued hydrologic processes. *ASCE Journal of Hydrologic Engineering* **6**(2): 109-118.

Saunders, M., P. Lewis and A. Thornhill (2003). *Research methods for business students*, 3rd ed., Financial Times Prentice Hall: London.

Schultz, C. L. (1987). Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society* **38**(5): 453-458.

Segerstedt, A. (2000). *Forecasting slow-moving items and ordinary items: a modification of Croston's idea*. Department of Business Administration and Social Science, Division of Industrial Logistics, Luleå University of Technology.

Shale, E. A., J. E. Boylan and F. R. Johnston (2006). Forecasting for intermittent demand: the estimation of an unbiased average. *Journal of the Operational Research Society* **57**(5): 588-592.

Shenstone, L. and R. J. Hyndman (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting* **24**(6): 389-402.

Silva, I. and M. E. Silva (2006). Asymptotic distribution of the Yule-Walker estimator for INAR(p) processes. *Statistics & Probability Letters* **76**(15): 1655-1663.

Silva, I., M. E. Silva, I. Pereira and N. Silva (2005). Replicated INAR(1) processes. *Methodology and Computing in Applied Probability* **7**(4): 517-542.

Silva, M. E. and V. L. Oliveira (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. *Journal of Time Series Analysis* **25**(3): 317-333.

Silva, M. E. and V. L. Oliveira (2005). Difference equations for the higher order moments and cumulants of the INAR(p) model. *Journal of Time Series Analysis* **26**(1): 17-36.

Silver, E. A., D. F. Pyke and R. Peterson (1998). *Inventory management and production planning and scheduling,* 3rd ed., John Wiley & Sons: New York.

Smith, J. (1979). A generalization of the Bayesian forecasting model. *Journal of the Royal Statistical Society. Series B* **41**(3): 375-387.

Snyder, R. (2002). Forecasting sales of slow and fast moving inventories. *European Journal of Operational Research* **140**(3): 684-699.

Snyder, R. D. and A. Beaumont (2007). *A comparison of methods for forecasting demand for slow moving car parts*. Monash Econometrics and Business Statistics Working Papers 15/07, Monash University.

Sprott, D. A. (1983). Estimating the parameters of a convolution by maximum likelihood. *Journal of the American Statistical Association* **78**(382): 457-460.

Stuart, A. and J. K. Ord (1994). *Kendall's advanced theory of statistics,* 6th Edition, Edward Arnold: London.

Sueutel, F. W. and K. van Harn (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**(5): 893-899.

Syntetos, A. A. (2001). *Forecasting for Intermittent Demand.* PhD thesis: Buckinghamshire Chilterns University College, Brunel University, UK.

Syntetos, A. A., M. Z. Babai, Y. Dallery and R. Teunter (2008). Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society*: doi:10.1057/palgrave.jors.2602593.

Syntetos, A. A. and J. E. Boylan (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics* **71**(1-3): 457–466.

Syntetos, A. A. and J. E. Boylan (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting* **21**(2): 303-314.

Syntetos, A. A., J. E. Boylan and J. D. Croston (2005). On the categorisation of demand patterns. *Journal of the Operational Research Society* **56**(5): 495-503.

Teles, P. and W. W. S. Wei (2002). The use of aggregate time series in testing for Gaussianity. *Journal of Time Series Analysis* **23**(1): 95-116.

Teunter, R. and B. Sani (2009). On the bias of Croston's forecasting method. *European Journal of Operational Research* **194**(1): 177-183.

Teunter, R. H. and L. Duncan (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society* **60**(3): 321-329.

Tiao, G. C. and D. Xu (1993). Robustness of maximum likelihood estimates for multi-step predictors: the exponential smoothing case. *Biometrika* **80**(3): 623-641.

Venkataraman, K. N. (1982). A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Advances in Applied Probability* **14**(1): 1-20.

Wald, A. and J. Wolfowitz (1940). On a test whether two samples are from the same population. *Annals of Mathematical Statistics* **11**(2): 147-162.

Weiß, C. H. (2008a). The combined INAR(p) models for time series of counts. *Statistics & Probability Letters* **78**(13): 1817-1822

Weiß, C. H. (2008b). Serial dependence and regression of Poisson INARMA models. *Journal of Statistical Planning and Inference* **138**(10): 2975-2990.

Weiß, C. H. (2008c). Thinning operations for modeling time series of counts-a survey. *AStA Advances in Statistical Analysis* **92**(3): 319-341.

Willemain, T. R., C. N. Smart and H. F. Schwarz (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting* **20**(3): 375-387.

Willemain, T. R., C. N. Smart, J. H. Shocker and P. A. DeSautels (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting* **10**(4): 529–538.

Williams, T. (1984). Stock control with sporadic and slow-moving demand. *Journal of the Operational Research Society* **35**(10): 939-948.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika* **75**(4): 621-629.

Zeger, S. L. and B. Qaqish (1988). Markov regression models for time series: a quasilikelihood approach. *Biometrics* **44**(4): 1019-1031.

Zheng, H., I. V. Basawa and S. Datta (2006). Inference for pth-order random coefficient integer-valued autoregressive processes. *Journal of Time Series Analysis* **27**(3): 411-440.

Zheng, H., I. V. Basawa and S. Datta (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference* **137**(1): 212-229

Zhu, R. and H. Joe (2006). Modelling count data time series with Markov processes based on Binomial thinning. *Journal of Time Series Analysis* **27**(5): 725–738.

Appendix 3.A Autocorrelation Function of an INARMA(1,1) Model

In this appendix, we show how to obtain the ACF of an INARMA(1,1) process of $Y_t = \alpha \circ Y_{t-1} + Z_t + \beta \circ Z_{t-1}$. The unconditional variance of the INARMA(1,1) process can be obtained from:

$$\mathrm{var}(Y_t) = \mathrm{var}(\alpha \circ Y_{t-1}) + \mathrm{var}(Z_t) + \mathrm{var}(\beta \circ Z_{t-1}) + 2\mathrm{cov}(\alpha \circ Y_{t-1}, \beta \circ Z_{t-1})$$

Considering the fact that $\mathrm{cov}(\alpha \circ X, \beta \circ Y) = \alpha\beta\mathrm{cov}(X, Y)$, the above equation can be written as:

$$\begin{aligned}\mathrm{var}(Y_t) &= \alpha^2\mathrm{var}(Y_{t-1}) + \alpha(1-\alpha)E(Y_{t-1}) + \sigma_Z^2 + \beta^2\sigma_Z^2 + \beta(1-\beta)\mu_Z \\ &\quad + 2\mathrm{cov}[(\alpha^2 \circ Y_{t-2} + \alpha \circ Z_{t-1} + \alpha\beta \circ Z_{t-2}), \beta \circ Z_{t-1}] \\ &= \alpha^2\mathrm{var}(Y_{t-1}) + \alpha(1-\alpha)\frac{\mu_Z(1+\beta)}{1-\alpha} + \sigma_Z^2 + \beta^2\sigma_Z^2 + \beta(1-\beta)\mu_Z \\ &\quad + 2\alpha\beta\sigma_Z^2\end{aligned}$$

Hence, the unconditional variance of an INARMA(1,1) process is:

$$\mathrm{var}(Y_t) = \frac{1}{1-\alpha^2}[(\alpha + \beta + \alpha\beta - \beta^2)\mu_Z + (1 + \beta^2 + 2\alpha\beta)\sigma_Z^2]$$

Equation 3.A-1

If a Poisson distribution is assumed for innovations ($\mu_Z = \sigma_Z^2 = \lambda$), the above result can be simplified to:

$$\mathrm{var}(Y_t) = \frac{\lambda}{1-\alpha^2}[1 + \alpha + \beta + 3\alpha\beta]$$

Equation 3.A-2

The INARMA(1,1) process, Y_t, can be written in terms of Y_{t-k} as follows:

$$Y_t = \alpha^k \circ Y_{t-k} + \sum_{j=0}^{k-1} \alpha^j \circ Z_{t-j} + \sum_{j=0}^{k-1} \alpha^j \beta \circ Z_{t-j-1}$$

Equation 3.A-3

The autocovariance at lag k is defined as:

$$\gamma_k = \mathrm{cov}(Y_t, Y_{t-k})$$

Equation 3.A-4

Therefore, the autocovariance at lag k is:

$$\begin{aligned}\gamma_k &= \operatorname{cov}\left[Y_{t-k}, \left(\alpha^k \circ Y_{t-k} + \sum_{j=0}^{k-1} \alpha^j \circ Z_{t-j} + \sum_{j=0}^{k-1} \alpha^j \beta \circ Z_{t-j-1}\right)\right] \\ &= \alpha^k \operatorname{var}(Y_{t-k}) + \operatorname{cov}\left(Y_{t-k}, \sum_{j=0}^{k-1} \alpha^j \circ Z_{t-j}\right) + \operatorname{cov}\left(Y_{t-k}, \sum_{j=0}^{k-1} \alpha^j \beta \circ Z_{t-j-1}\right) \\ &= \alpha^k \operatorname{var}(Y_{t-k}) + \operatorname{cov}(Y_{t-k}, \alpha^{k-1}\beta \circ Z_{t-k}) \\ &= \alpha^k \operatorname{var}(Y_{t-k}) + \operatorname{cov}(\alpha \circ Y_{t-k-1} + Z_{t-k} + \beta \circ Z_{t-k-1}, \alpha^{k-1}\beta \circ Z_{t-k})\end{aligned}$$

By substitution from Equation 3.A-1:

$$\begin{aligned}\gamma_k &= \frac{\alpha^k}{1-\alpha^2}\left[(\alpha + \beta + \alpha\beta - \beta^2)\mu_Z + (1 + \beta^2 + 2\alpha\beta)\sigma_Z^2\right] + \alpha^{k-1}\beta\sigma_Z^2 \\ &= \alpha^{k-1}\left\{\left[\frac{\alpha(\alpha + \beta + \alpha\beta - \beta^2)}{1-\alpha^2}\right]\mu_Z + \left[\frac{\alpha(1 + \beta^2 + 2\alpha\beta)}{1-\alpha^2} + \beta\right]\sigma_Z^2\right\}\end{aligned}$$

As a result:

$$\gamma_k = \alpha^{k-1}\left\{\left[\frac{\alpha^2 + \alpha\beta + \alpha^2\beta - \alpha\beta^2}{1-\alpha^2}\right]\mu_Z + \left[\frac{\alpha + \alpha\beta^2 + \alpha^2\beta + \beta}{1-\alpha^2}\right]\sigma_Z^2\right\}$$

Equation 3.A-5

Consequently, the ACF of an INARMA(1,1) is:

$$\rho_k = \begin{cases} \dfrac{(\alpha^2 + \alpha\beta + \alpha^2\beta - \alpha\beta^2)\mu_Z + (\alpha + \alpha\beta^2 + \alpha^2\beta + \beta)\sigma_Z^2}{(\alpha + \beta + \alpha\beta - \beta^2)\mu_Z + (1 + \beta^2 + 2\alpha\beta)\sigma_Z^2} & \text{for } k = 1 \\ \alpha\rho_{k-1} & \text{for } k > 1 \end{cases}$$

Equation 3.A-6

For a PoINARMA(1,1) with $\mu_Z = \sigma_Z^2 = \lambda$, the ACF would be:

$$\rho_k = \begin{cases} \dfrac{\alpha + \beta + \alpha\beta + \alpha^2 + 2\alpha^2\beta}{1 + \alpha + \beta + 3\alpha\beta} & \text{for } k = 1 \\ \alpha\rho_{k-1} & \text{for } k > 1 \end{cases}$$

Equation 3.A-7

Appendix 3.B The Unconditional Variance of an INARMA(*p*,*q*) Model

The unconditional variance of an INARMA(p,q) process of $Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i \circ Z_{t-i}$, can be written as follows, where all the covariance terms, to be found later, have been summarized using the expression "(cov) terms":

$$\begin{aligned}\mathrm{var}(Y_t) &= \mathrm{var}(\alpha_1 \circ Y_{t-1}) + \cdots + \mathrm{var}(\alpha_p \circ Y_{t-p}) + \mathrm{var}(Z_t) \\ &\quad + \mathrm{var}(\beta_1 \circ Z_{t-1}) + \cdots + \mathrm{var}(\beta_q \circ Z_{t-q}) + (\mathrm{cov})\ \mathrm{terms}\end{aligned}$$

Equation 3.B-1

$$\begin{aligned}\mathrm{var}(Y_t) &= \alpha_1^2 \mathrm{var}(Y_{t-1}) + \alpha_1(1-\alpha_1)E(Y_{t-1}) + \cdots + \alpha_p^2 \mathrm{var}(Y_{t-p}) \\ &\quad + \alpha_p(1-\alpha_p)E(Y_{t-p}) + \sigma_Z^2 + \beta_1^2\sigma_Z^2 + \beta_1(1-\beta_1)\mu_Z + \cdots + \beta_q^2\sigma_Z^2 \\ &\quad + \beta_q(1-\beta_q)\mu_Z + (\mathrm{cov})\ \mathrm{terms}\end{aligned}$$

Equation 3.B-2

Hence, assuming stationarity of the process, we have:

$$\begin{aligned}\mathrm{var}(Y_t) &= \frac{1}{1-\sum_{i=1}^{p}\alpha_i^2}\left\{\left[\frac{1+\sum_{i=1}^{q}\beta_i}{1-\sum_{i=1}^{p}\alpha_i}\sum_{i=1}^{p}\alpha_i(1-\alpha_i) + \sum_{i=1}^{q}\beta_i(1-\beta_i)\right]\mu_Z \right. \\ &\quad \left. + \left[1+\sum_{i=1}^{q}\beta_i^2\right]\sigma_Z^2 + (\mathrm{cov})\ \mathrm{terms}\right\}\end{aligned}$$

Equation 3.B-3

Next, we focus on finding the covariance terms. There are two types of covariance:

1. The covariance between $\{Y_t\}$s at different lags: $\mathrm{cov}(Y_i, Y_j)$ for $i \neq j$.

2. The covariance between $\{Y_t\}$s and $\{Z_t\}$s: $\mathrm{cov}(Y_i, Z_j)$ for $i \geq j$.

   For $j > i$, the random disturbance terms are independent of previous observations.

The first group of covariance terms is given by:

$$\begin{aligned}\text{first Cov} = {} & 2\text{cov}(\alpha_1 \circ Y_{t-1}, \alpha_2 \circ Y_{t-2}) + 2\text{cov}(\alpha_1 \circ Y_{t-1}, \alpha_3 \circ Y_{t-3}) + \cdots \\ & +2\text{cov}\left(\alpha_1 \circ Y_{t-1}, \alpha_p \circ Y_{t-p}\right) \\ & +2\text{cov}(\alpha_2 \circ Y_{t-2}, \alpha_3 \circ Y_{t-3}) + 2\text{cov}(\alpha_2 \circ Y_{t-2}, \alpha_4 \circ Y_{t-4}) + \cdots \\ & + 2\text{cov}\left(\alpha_2 \circ Y_{t-2}, \alpha_p \circ Y_{t-p}\right) + \cdots \\ & + 2\text{cov}\left(\alpha_{p-2} \circ Y_{t-p+2}, \alpha_{p-1} \circ Y_{t-p+1}\right) + 2\text{cov}\left(\alpha_{p-2} \circ Y_{t-p+2}, \alpha_p \circ Y_{t-p}\right) \\ & + 2\text{cov}\left(\alpha_{p-1} \circ Y_{t-p+1}, \alpha_p \circ Y_{t-p}\right)\end{aligned}$$

$$\begin{aligned}\text{first Cov} = {} & 2\alpha_1\alpha_2\gamma_1 + 2\alpha_1\alpha_3\gamma_2 + \cdots + 2\alpha_1\alpha_p\gamma_{p-1} \\ & +2\alpha_2\alpha_3\gamma_1 + 2\alpha_2\alpha_4\gamma_2 + \cdots + 2\alpha_2\alpha_p\gamma_{p-2} + \cdots \\ & +2\,\alpha_{p-2}\alpha_{p-1}\gamma_1 + 2\alpha_{p-2}\alpha_p\gamma_2 \\ & +2\alpha_{p-1}\alpha_p\gamma_1\end{aligned}$$

$$\text{first Cov} = 2\sum_{i=1}^{p-1} \alpha_i\alpha_{i+1}\gamma_1 + 2\sum_{i=1}^{p-2} \alpha_i\alpha_{i+2}\gamma_2 + \cdots + 2\sum_{i=1}^{p-(p-1)} \alpha_i\alpha_{i+(p-1)}\gamma_{p-1}$$

$$\text{first Cov} = 2\sum_{j=1}^{p-1}\sum_{i=1}^{p-j} \alpha_i\alpha_{i+j}\gamma_j$$

Equation 3.B-4

Now we focus on the second type of covariance which is $\text{cov}(Y_i, Z_j)$ for $i \geq j$. Here there are three cases:

1. $p < q$
2. $p = q$
3. $p > q$

Each of these cases is discussed here separately.

1. $\mathrm{cov}(Y_i, Z_j)$ when $p < q$

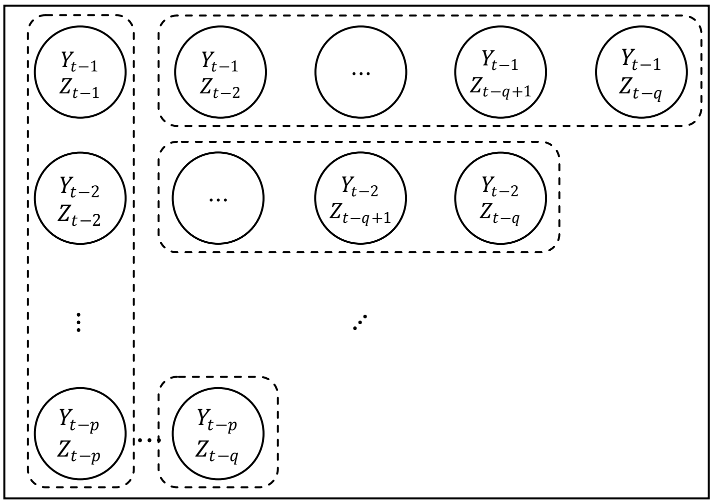

Figure 3.B-1 $\mathrm{cov}(Y_i, Z_j)$ in an INARMA(p,q) process when $p < q$

$$\begin{aligned}
\text{Second Cov 1} = {} & 2\mathrm{cov}(\alpha_1 \circ Y_{t-1}, \beta_1 \circ Z_{t-1}) + 2\mathrm{cov}(\alpha_1 \circ Y_{t-1}, \beta_2 \circ Z_{t-2}) + \cdots \\
& + 2\mathrm{cov}\left(\alpha_1 \circ Y_{t-1}, \beta_q \circ Z_{t-q}\right) \\
& + 2\mathrm{cov}(\alpha_2 \circ Y_{t-2}, \beta_2 \circ Z_{t-2}) + 2\mathrm{cov}(\alpha_2 \circ Y_{t-2}, \beta_3 \circ Z_{t-3}) + \cdots \\
& + 2\mathrm{cov}\left(\alpha_2 \circ Y_{t-2}, \beta_q \circ Z_{t-q}\right) + \cdots \\
& + 2\mathrm{cov}\left(\alpha_p \circ Y_{t-p}, \beta_p \circ Z_{t-p}\right) + \cdots + 2\mathrm{cov}\left(\alpha_p \circ Y_{t-p}, \beta_q \circ Z_{t-q}\right)
\end{aligned}$$

Then, the first terms in each of the rows are summed vertically, with the remaining terms in the rows being summed horizontally (see Figure 3.B-1).

$$\begin{aligned}
\text{Second Cov 1} = {} & 2\sigma_Z^2 \sum_{i=1}^{p} \alpha_i \beta_i + 2\sum_{j=2}^{q} \alpha_1 \beta_j \gamma_{j-1}^{YZ} + 2\sum_{j=3}^{q} \alpha_2 \beta_j \gamma_{j-2}^{YZ} \\
& + \cdots + 2\sum_{j=p+1}^{q} \alpha_p \beta_j \gamma_{j-p}^{YZ}
\end{aligned}$$

$$\text{Second Cov 1} = 2\sigma_Z^2 \sum_{i=1}^{p} \alpha_i \beta_i + 2\sum_{i=1}^{p} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ}$$

Equation 3.B-5

where γ_k^{YZ} is the cross-covariance between Y and Z at lag k [$\gamma_k^{YZ} = E(Y_t Z_{t-k})$] and is analysed further in appendix 3.C.

2. $\text{cov}(Y_i, Z_j)$ when $p = q$

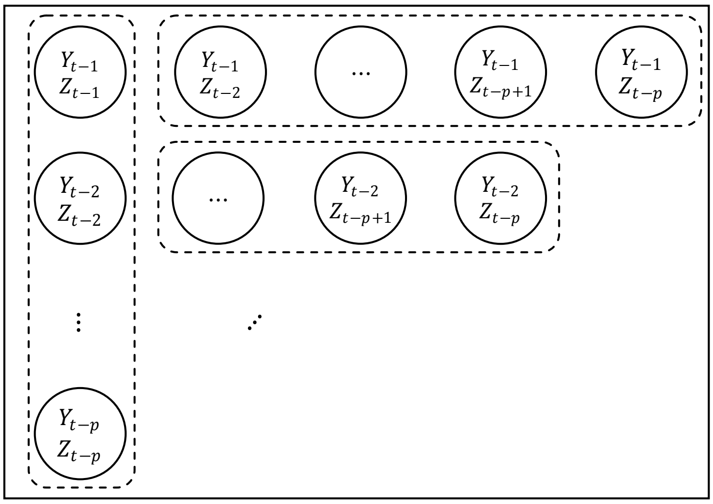

Figure 3.B-2 $\text{cov}(Y_i, Z_j)$ in an INARMA(p,q) process when $p = q$

$$\begin{aligned}
\text{Second Cov 2} = {} & 2\text{cov}(\alpha_1 \circ Y_{t-1}, \beta_1 \circ Z_{t-1}) + 2\text{cov}(\alpha_1 \circ Y_{t-1}, \beta_2 \circ Z_{t-2}) + \cdots \\
& +2\text{cov}\left(\alpha_1 \circ Y_{t-1}, \beta_p \circ Z_{t-p}\right) \\
& +2\text{cov}(\alpha_2 \circ Y_{t-2}, \beta_2 \circ Z_{t-2}) + 2\text{cov}(\alpha_2 \circ Y_{t-2}, \beta_3 \circ Z_{t-3}) + \cdots \\
& + 2\text{cov}\left(\alpha_2 \circ Y_{t-2}, \beta_p \circ Z_{t-p}\right) + \cdots \\
& +2\text{cov}\left(\alpha_{p-1} \circ Y_{t-p+1}, \beta_{p-1} \circ Z_{t-p+1}\right) \\
& + 2\text{cov}\left(\alpha_{p-1} \circ Y_{t-p+1}, \beta_p \circ Z_{t-p}\right) \\
& +2\text{cov}\left(\alpha_p \circ Y_{t-p}, \beta_p \circ Z_{t-p}\right)
\end{aligned}$$

Then, the first terms in each of the rows are summed vertically, with the remaining terms in the rows being summed horizontally (see Figure 3.B-2).

$$\begin{aligned}
\text{Second Cov 2} = {} & 2\sigma_Z^2 \sum_{i=1}^{p} \alpha_i \beta_i + 2\sum_{j=2}^{p} \alpha_1 \beta_j \gamma_{j-1}^{YZ} + 2\sum_{j=3}^{p} \alpha_2 \beta_j \gamma_{j-2}^{YZ} + \cdots \\
& +2\sum_{j=p}^{p} \alpha_{p-1} \beta_j \gamma_{j-(p-1)}^{YZ}
\end{aligned}$$

$$\text{Second Cov 2} = 2\sigma_Z^2 \sum_{i=1}^{p} \alpha_i \beta_i + 2\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \alpha_i \beta_j \gamma_{j-i}^{YZ}$$

Equation 3.B-6

The Equation 3.B-6 can be written as $2\sigma_Z^2 \sum_{i=1}^{p} \alpha_i \beta_i + 2\sum_{i=1}^{p} \sum_{j=i+1}^{p} \alpha_i \beta_j \gamma_{j-i}^{YZ}$ for generalization purposes because for $i = p$ the second summation would be $(\sum_{j=p+1}^{p})$ which has zero terms.

3. $\text{cov}(Y_i, Z_j)$ when $p > q$

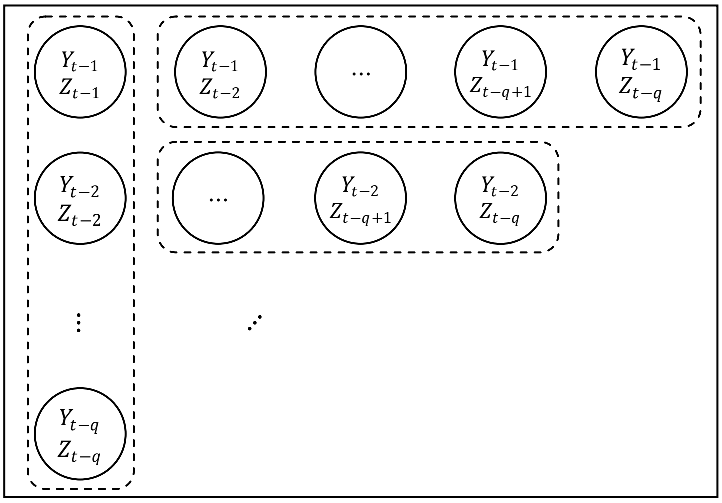

Figure 3.B-3 $\text{cov}(Y_i, Z_j)$ in an INARMA(p,q) process when $p > q$

$$\begin{aligned}\text{Second Cov 3} = {} & 2\text{cov}(\alpha_1 \circ Y_{t-1}, \beta_1 \circ Z_{t-1}) + 2\text{cov}(\alpha_1 \circ Y_{t-1}, \beta_2 \circ Z_{t-2}) + \cdots \\ & +2\text{cov}\left(\alpha_1 \circ Y_{t-1}, \beta_q \circ Z_{t-q}\right) \\ & +2\text{cov}(\alpha_2 \circ Y_{t-2}, \beta_2 \circ Z_{t-2}) + 2\text{cov}(\alpha_2 \circ Y_{t-2}, \beta_3 \circ Z_{t-3}) + \cdots \\ & + 2\text{cov}\left(\alpha_2 \circ Y_{t-2}, \beta_q \circ Z_{t-q}\right) + \cdots \\ & +2\text{cov}\left(\alpha_{q-1} \circ Y_{t-q+1}, \beta_{q-1} \circ Z_{t-q+1}\right) \\ & + 2\text{cov}\left(\alpha_{q-1} \circ Y_{t-q+1}, \beta_q \circ Z_{t-q}\right) \\ & +2\text{cov}\left(\alpha_q \circ Y_{t-q}, \beta_q \circ Z_{t-q}\right)\end{aligned}$$

Then, the first terms in each of the rows are summed vertically, with the remaining terms in the rows being summed horizontally (see Figure 3.B-3).

$$\begin{aligned}\text{Second Cov 3} = {} & 2\sigma_Z^2 \sum_{i=1}^{q} \alpha_i \beta_i + 2\sum_{j=2}^{q} \alpha_1 \beta_j \gamma_{j-1}^{YZ} + 2\sum_{j=3}^{q} \alpha_2 \beta_j \gamma_{j-2}^{YZ} + \cdots \\ & +2\sum_{j=q}^{q} \alpha_{q-1} \beta_j \gamma_{j-(q-1)}^{YZ}\end{aligned}$$

$$\text{Second Cov 3} = 2\sigma_Z^2 \sum_{i=1}^{q} \alpha_i \beta_i + 2 \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ}$$

Equation 3.B-7

The Equation 3.B-7 can be written as $2\sigma_Z^2 \sum_{i=1}^{q} \alpha_i \beta_i + 2 \sum_{i=1}^{q} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ}$ for generalization purposes because for $i = q$ the second summation would be $(\sum_{j=q+1}^{q})$ which has zero terms.

Therefore, based on the results of Equation 3.B-5, Equation 3.B-6 and Equation 3.B-7, the second group of covariance is given by:

$$\text{Second Cov} = 2 \sum_{i=1}^{\min(p,q)} \alpha_i \beta_i \sigma_Z^2 + 2 \sum_{i=1}^{\min(p,q)} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ}$$

Equation 3.B-8

Finally, from the Equation 3.B-3, Equation 3.B-4, and Equation 3.B-8, the unconditional variance of an INARMA(p,q) process is:

$$\begin{aligned}\text{var}(Y_t) = \frac{1}{1-\sum_{i=1}^{p} \alpha_i^2} \Bigg\{ &\left[\frac{1+\sum_{i=1}^{q} \beta_i}{1-\sum_{i=1}^{p} \alpha_i} \sum_{i=1}^{p} \alpha_i (1-\alpha_i) + \sum_{i=1}^{q} \beta_i (1-\beta_i) \right] \mu_Z \\ &+ \left[1 + \sum_{i=1}^{q} \beta_i^2 \right] \sigma_Z^2 + 2 \sum_{j=1}^{p-1} \sum_{i=1}^{p-j} \alpha_i \alpha_{i+j} \gamma_j \\ &+ 2 \sum_{i=1}^{\min(p,q)} \alpha_i \beta_i \sigma_Z^2 + 2 \sum_{i=1}^{\min(p,q)} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ} \Bigg\}\end{aligned}$$

$$\begin{aligned}\text{var}(Y_t) = &\frac{\mu_Z}{1-\sum_{i=1}^{p} \alpha_i^2} \left[\frac{1+\sum_{i=1}^{q} \beta_i}{1-\sum_{i=1}^{p} \alpha_i} \sum_{i=1}^{p} \alpha_i (1-\alpha_i) + \sum_{i=1}^{q} \beta_i (1-\beta_i) \right] \\ &+ \frac{\sigma_Z^2}{1-\sum_{i=1}^{p} \alpha_i^2} \left[1 + \sum_{i=1}^{q} \beta_i^2 + 2 \sum_{i=1}^{\min(p,q)} \alpha_i \beta_i \right] \\ &+ \frac{2 \sum_{j=1}^{p-1} \sum_{i=1}^{p-j} \alpha_i \alpha_{i+j} \gamma_j + 2 \sum_{i=1}^{\min(p,q)} \sum_{j=i+1}^{q} \alpha_i \beta_j \gamma_{j-i}^{YZ}}{1-\sum_{i=1}^{p} \alpha_i^2}\end{aligned}$$

Equation 3.B-9

Appendix 3.C The Cross-Covariance Function between Y and Z for an INARMA(p,q) Model

The cross-covariance function, γ_k^{YZ}, is the covariance between Y and Z at lag k and is defined by $\gamma_k^{YZ} = E(Y_t Z_{t-k})$. Therefore, the cross-covariance at lag zero is given by:

$$\gamma_0^{YZ} = \mathrm{cov}(Y_t, Z_t) = \mathrm{cov}\left(\sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + Z_t + \sum_{i=1}^{q} \beta_i \circ Z_{t-i}, Z_t\right)$$

Considering the fact that the innovation terms are independent of previous observations, we have:

$$\gamma_0^{YZ} = \mathrm{cov}(Z_t, Z_t) = var(Z_t) = \sigma_Z^2$$

The cross-covariance at lag one can be obtained from:

$$\begin{aligned}
\gamma_1^{YZ} &= \mathrm{cov}(Y_t, Z_{t-1}) \\
&= \mathrm{cov}\left(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}, Z_{t-1}\right) \\
&= \mathrm{cov}(\alpha_1 \circ Y_{t-1}, Z_{t-1}) + \mathrm{cov}(\beta_1 \circ Z_{t-1}, Z_{t-1}) = \alpha_1 \sigma_Z^2 + \beta_1 \sigma_Z^2
\end{aligned}$$

The cross-covariance at lag k $(0 \leq k \leq q)$ is given by:

$$\begin{aligned}
\gamma_k^{YZ} &= \mathrm{cov}(Y_t, Z_{t-k}) \\
&= \mathrm{cov}\left(\alpha_1 \circ Y_{t-1} + \cdots + \alpha_p \circ Y_{t-p} + Z_t + \beta_1 \circ Z_{t-1} + \cdots + \beta_q \circ Z_{t-q}, Z_{t-k}\right) \\
&= \mathrm{cov}(\alpha_1 \circ Y_{t-1}, Z_{t-k}) + \cdots + \mathrm{cov}\left(\alpha_{k-1} \circ Y_{t-(k-1)}, Z_{t-k}\right) \\
&\quad + \mathrm{cov}(\alpha_k \circ Y_{t-k}, Z_{t-k}) + \beta_k \sigma_Z^2 \\
&= \beta_k \sigma_Z^2 + \alpha_1 \gamma_{k-1}^{YZ} + \cdots + \alpha_{k-1} \gamma_1^{YZ} + \alpha_k \sigma_Z^2
\end{aligned}$$

$$\gamma_k^{YZ} = \beta_k \sigma_Z^2 + \sum_{i=1}^{k} \alpha_i \gamma_{k-i}^{YZ}$$

Equation 3.C-1

Appendix 3.D Over Lead Time Aggregation of an INAR(1) Model

In this appendix, we show how to derive the conditional first and second moments of a lead time aggregated PoINAR(1) process ($\mu_Z = \sigma_Z^2 = \lambda$). The aggregated process over lead time can be written as:

$$\begin{aligned}
\sum_{i=1}^{l+1} Y_{t+i} &= (\alpha \circ Y_t + Z_{t+1}) + (\alpha \circ Y_{t+1} + Z_{t+2}) + \cdots + (\alpha \circ Y_{t+l} + Z_{t+l+1}) \\
&= (\alpha \circ Y_t + Z_{t+1}) + (\alpha \circ (\alpha \circ Y_t + Z_{t+1}) + Z_{t+2}) + \cdots + (\alpha \circ Y_{t+l} + Z_{t+l+1}) \\
&= (\alpha \circ Y_t + Z_{t+1}) + (\alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + Z_{t+2}) + \cdots \\
&+ (\alpha^{l+1} \circ Y_t + \alpha^l \circ Z_{t+1} + \alpha^{l-1} \circ Z_{t+2} + \cdots + \alpha \circ Z_{t+l} + Z_{t+l+1})
\end{aligned}$$

It can be simplified as

$$\begin{aligned}
\sum_{i=1}^{l+1} Y_{t+i} &= (\alpha \circ Y_t + \alpha^2 \circ Y_t + \cdots + \alpha^{l+1} \circ Y_t) \\
&+ (Z_{t+1} + \alpha \circ Z_{t+1} + \cdots + \alpha^l \circ Z_{t+1}) \\
&+ (Z_{t+2} + \alpha \circ Z_{t+2} + \cdots + \alpha^{l-1} \circ Z_{t+2}) + \cdots + (Z_{t+l} + \alpha \circ Z_{t+l}) + Z_{t+l+1}
\end{aligned}$$

Equation 3.D-1

The conditional expected value of the aggregated process is given by:

$$\begin{aligned}
E\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] &= (\alpha Y_t + \alpha^2 Y_t + \cdots + \alpha^{l+1} Y_t) + (\lambda + \alpha\lambda + \cdots + \alpha^l \lambda) \\
&+ (\lambda + \alpha\lambda + \cdots + \alpha^{l-1}\lambda) + \cdots + (\lambda + \alpha\lambda) + \lambda
\end{aligned}$$

$$E\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\lambda}{1-\alpha}[(1-\alpha^{l+1}) + (1-\alpha^l) + \cdots + (1-\alpha)]$$

Therefore, we have:

$$E\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$

Equation 3.D-2

Now we want to find the conditional variance of the aggregated process. We know that:

$$\operatorname{cov}\left(\alpha^i \circ X, \alpha^j \circ X\right) = \alpha^i \alpha^j E(X^2) - \alpha^i E(X) \alpha^j E(X) = \alpha^i \alpha^j \operatorname{var}(X)$$

Equation 3.D-3

Hence, we have $\operatorname{cov}\left(\alpha^i \circ Z_t, \alpha^j \circ Z_t\right) = \alpha^i \alpha^j \lambda$.

The variance of the Equation 3.D-1 given Y_t is:

$$\begin{aligned}
\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] &= \operatorname{var}(\alpha \circ Y_t) + \operatorname{var}(\alpha^2 \circ Y_t) + \cdots + \operatorname{var}\left(\alpha^{l+1} \circ Y_t\right) \\
&+2\operatorname{cov}(\alpha \circ Y_t, \alpha^2 \circ Y_t) + 2\operatorname{cov}(\alpha \circ Y_t, \alpha^3 \circ Y_t) + \cdots + 2\operatorname{cov}\left(\alpha \circ Y_t, \alpha^{l+1} \circ Y_t\right) \\
&+ 2\operatorname{cov}(\alpha^2 \circ Y_t, \alpha^3 \circ Y_t) + 2\operatorname{cov}(\alpha^2 \circ Y_t, \alpha^4 \circ Y_t) + \cdots + 2\operatorname{cov}\left(\alpha^2 \circ Y_t, \alpha^{l+1} \circ Y_t\right) \\
&+ \cdots + 2\operatorname{cov}\left(\alpha^{l-1} \circ Y_t, \alpha^l \circ Y_t\right) + 2\operatorname{cov}\left(\alpha^{l-1} \circ Y_t, \alpha^{l+1} \circ Y_t\right) \\
&+2\operatorname{cov}\left(\alpha^l \circ Y_t, \alpha^{l+1} \circ Y_t\right) \\
&+\operatorname{var}(Z_{t+1}) + \operatorname{var}(\alpha \circ Z_{t+1}) + \operatorname{var}(\alpha^2 \circ Z_{t+1}) + \cdots + \operatorname{var}\left(\alpha^l \circ Z_{t+1}\right) \\
&+2\operatorname{cov}(Z_{t+1}, \alpha \circ Z_{t+1}) + 2\operatorname{cov}(Z_{t+1}, \alpha^2 \circ Z_{t+1}) + \cdots + 2\operatorname{cov}\left(Z_{t+1}, \alpha^l \circ Z_{t+1}\right) \\
&+2\operatorname{cov}(\alpha \circ Z_{t+1}, \alpha^2 \circ Z_{t+1}) + 2\operatorname{cov}(\alpha \circ Z_{t+1}, \alpha^3 \circ Z_{t+1}) + \cdots \\
&+ 2\operatorname{cov}\left(\alpha \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right) \\
&+ \cdots + 2\operatorname{cov}\left(\alpha^{l-2} \circ Z_{t+1}, \alpha^{l-1} \circ Z_{t+1}\right) + 2\operatorname{cov}\left(\alpha^{l-2} \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right) \\
&+2\operatorname{cov}\left(\alpha^{l-1} \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right) \\
&+\operatorname{var}(Z_{t+2}) + \operatorname{var}(\alpha \circ Z_{t+2}) + \cdots + \operatorname{var}\left(\alpha^{l-1} \circ Z_{t+2}\right) \\
&+2\operatorname{cov}(Z_{t+2}, \alpha \circ Z_{t+2}) + 2\operatorname{cov}(Z_{t+2}, \alpha^2 \circ Z_{t+2}) + \cdots + 2\operatorname{cov}\left(Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right) \\
&+2\operatorname{cov}(\alpha \circ Z_{t+2}, \alpha^2 \circ Z_{t+2}) + 2\operatorname{cov}(\alpha \circ Z_{t+2}, \alpha^3 \circ Z_{t+2}) + \cdots \\
&+ 2\operatorname{cov}\left(\alpha \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right) \\
&+ \cdots + 2\operatorname{cov}\left(\alpha^{l-3} \circ Z_{t+2}, \alpha^{l-2} \circ Z_{t+2}\right) + 2\operatorname{cov}\left(\alpha^{l-3} \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right) \\
&+2\operatorname{cov}\left(\alpha^{l-2} \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right) + \cdots \\
&+\operatorname{var}(Z_{t+l-1}) + \operatorname{var}(\alpha \circ Z_{t+l-1}) + \operatorname{var}(\alpha^2 \circ Z_{t+l-1}) \\
&+2\operatorname{cov}(Z_{t+l-1}, \alpha \circ Z_{t+l-1}) + 2\operatorname{cov}(Z_{t+l-1}, \alpha^2 \circ Z_{t+l-1}) \\
&+ 2\operatorname{cov}(\alpha \circ Z_{t+l-1}, \alpha^2 \circ Z_{t+l-1}) \\
&+\operatorname{var}(Z_{t+l}) + \operatorname{var}(\alpha \circ Z_{t+l}) \\
&+2\operatorname{cov}(Z_{t+l}, \alpha \circ Z_{t+l}) \\
&+\operatorname{var}(Z_{t+l+1})
\end{aligned}$$

Since Y_t is fixed, $\operatorname{cov}\left(\alpha^i \circ Y_t, \alpha^j \circ Y_t\right) = \alpha^i \alpha^j \operatorname{var}(Y_t) = 0$, using Equation 3.D-3.

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = \alpha(1-\alpha)E(Y_t) + \alpha^2(1-\alpha^2)E(Y_t) + \cdots + \alpha^{l+1}(1-\alpha^{l+1})E(Y_t)$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda] + \cdots + \left[\alpha^{2l}\lambda + \alpha^l\left(1-\alpha^l\right)\lambda\right]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^l]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^{l+1}]\lambda + \cdots + 2[\alpha^{2l-3} + \alpha^{2l-2}]\lambda$$
$$+ 2[\alpha^{2l-1}]\lambda$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda] + \cdots$$
$$+ \left[\alpha^{2l-2}\lambda + \alpha^{l-1}\left(1-\alpha^{l-1}\right)\lambda\right]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^{l-1}]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^l]\lambda + \cdots + 2[\alpha^{2l-5} + \alpha^{2l-4}]\lambda$$
$$+ 2[\alpha^{2l-3}]\lambda + \cdots$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda]$$
$$+2[\alpha + \alpha^2]\lambda + 2[\alpha^3]\lambda$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda]$$
$$+2\alpha\lambda$$
$$+\lambda$$

The above result can be summarized to:

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = \alpha(1-\alpha)Y_t + \alpha^2(1-\alpha^2)Y_t + \cdots + \alpha^{l+1}(1-\alpha^{l+1})Y_t$$
$$+\lambda + [\alpha\lambda] + [\alpha^2\lambda] + \cdots + [\alpha^l\lambda]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^l]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^{l+1}]\lambda + \cdots + 2[\alpha^{2l-3} + \alpha^{2l-2}]\lambda$$
$$+ 2[\alpha^{2l-1}]\lambda$$
$$+\lambda + [\alpha\lambda] + [\alpha^2\lambda] + \cdots + [\alpha^{l-1}\lambda]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^{l-1}]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^l]\lambda + \cdots + 2[\alpha^{2l-5} + \alpha^{2l-4}]\lambda$$
$$+ 2[\alpha^{2l-3}]\lambda + \cdots$$
$$+\lambda + [\alpha\lambda] + [\alpha^2\lambda]$$
$$+2[\alpha + \alpha^2]\lambda + 2[\alpha^3]\lambda$$
$$+\lambda + [\alpha\lambda]$$
$$+2\alpha\lambda$$
$$+\lambda$$

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i}\,|Y_t\right] = \alpha(1-\alpha)Y_t + \alpha^2(1-\alpha^2)Y_t + \cdots + \alpha^{l+1}(1-\alpha^{l+1})Y_t$$
$$+\lambda\left(1 + \alpha + \cdots + \alpha^l\right) + \lambda\left(1 + \alpha + \cdots + \alpha^{l-1}\right) + \cdots + \lambda(1 + \alpha + \alpha^2) + \lambda(1+\alpha)$$
$$+ \lambda$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^l]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^{l+1}]\lambda + \cdots + 2[\alpha^{2l-3} + \alpha^{2l-2}]\lambda$$
$$+ 2[\alpha^{2l-1}]\lambda$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^{l-1}]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^l]\lambda + \cdots + 2[\alpha^{2l-5} + \alpha^{2l-4}]\lambda$$
$$+ 2[\alpha^{2l-3}]\lambda + \cdots$$
$$+2[\alpha + \alpha^2]\lambda + 2[\alpha^3]\lambda$$
$$+2\alpha\lambda$$

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = Y_t \sum_{j=1}^{l+1} \alpha^j \left(1-\alpha^j\right) + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$
$$+2\alpha\lambda\left[1+\alpha+\cdots+\alpha^{l-1}\right] + 2\alpha^3\lambda\left[1+\alpha+\cdots+\alpha^{l-2}\right] + \cdots$$
$$+2\alpha^{2l-3}\lambda[1+\alpha] + 2\alpha^{2l-1}\lambda$$
$$+2\alpha\lambda\left[1+\alpha+\cdots+\alpha^{l-2}\right] + 2\alpha^3\lambda\left[1+\alpha+\cdots+\alpha^{l-3}\right] + \cdots$$
$$+2\alpha^{2l-5}\lambda[1+\alpha] + 2\alpha^{2l-3}\lambda + \cdots$$
$$+2\alpha\lambda[1+\alpha] + 2\alpha^3\lambda$$
$$+2\alpha\lambda$$

So

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = Y_t \sum_{j=1}^{l+1} \alpha^j \left(1-\alpha^j\right) + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$
$$+\frac{2\alpha\lambda}{1-\alpha}\left[1-\alpha^l\right] + \frac{2\alpha^3\lambda}{1-\alpha}\left[1-\alpha^{l-1}\right] + \cdots + \frac{2\alpha^{2l-3}\lambda}{1-\alpha}\left[1-\alpha^2\right] + \frac{2\alpha^{2l-1}\lambda}{1-\alpha}[1-\alpha]$$
$$+\frac{2\alpha\lambda}{1-\alpha}\left[1-\alpha^{l-1}\right] + \frac{2\alpha^3\lambda}{1-\alpha}\left[1-\alpha^{l-2}\right] + \cdots + \frac{2\alpha^{2l-5}\lambda}{1-\alpha}\left[1-\alpha^2\right]$$
$$+\frac{2\alpha^{2l-3}\lambda}{1-\alpha}[1-\alpha]$$
$$+\cdots$$
$$+\frac{2\alpha\lambda}{1-\alpha}\left[1-\alpha^2\right] + \frac{2\alpha^3\lambda}{1-\alpha}[1-\alpha]$$
$$+\frac{2\alpha\lambda}{1-\alpha}[1-\alpha]$$

Summing the above expressions vertically results in:

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = Y_t \sum_{j=1}^{l+1} \alpha^j \left(1-\alpha^j\right) + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$
$$+\frac{2\alpha\lambda}{1-\alpha}\left[l-(\alpha+\alpha^2+\cdots+\alpha^l)\right] + \frac{2\alpha^3\lambda}{1-\alpha}\left[(l-1)-(\alpha+\alpha^2+\cdots+\alpha^{l-1})\right]$$
$$+\cdots+\frac{2\alpha^{2l-3}\lambda}{1-\alpha}\left[2-(\alpha+\alpha^2)\right] + \frac{2\alpha^{2l-1}\lambda}{1-\alpha}[1-\alpha]$$

Finally, the conditional variance of the aggregated process is:

$$\operatorname{var}\left[\sum_{i=1}^{l+1} Y_{t+i} \,|Y_t\right] = Y_t \sum_{j=1}^{l+1} \alpha^j \left(1-\alpha^j\right) + \frac{\lambda}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right]$$

$$+ \frac{2\lambda}{1-\alpha} \sum_{j=1}^{l} \alpha^{2j-1} \left[(l-j+1) - \frac{\alpha(1-\alpha^{l-j+1})}{1-\alpha}\right]$$

Equation 3.D-4

Appendix 4.A Infinite Autoregressive Representation of an INARMA(p,q) Model

It will be shown in this appendix that the Infinite Auto-Regressive Representation (IARR) of an INARMA(p,q) process ($Y_t = \sum_{j=1}^{p} \alpha_j \circ Y_{t-j} + Z_t + \sum_{j=1}^{q} \beta_j \circ Z_{t-j}$) is as follows:

$$Y_t = Z_t + \sum_{j=1}^{\infty} \sum_{i=1}^{n_j} \psi_{ij} \circ Y_{t-j}$$

Equation 4.A-1

where

$$n_j = \begin{cases} \left(\sum_{i=1}^{q} n_{j-i} \right) + 1 & 0 < j \leq p \\ \sum_{i=1}^{q} n_{j-i} & j > p \end{cases}$$

Equation 4.A-2

Then Z_t can be expressed in terms of $\{Y_{t-j}\}_{j=0}^{\infty}$ as:

$$Z_t = Y_t - \sum_{j=1}^{\infty} \sum_{i=1}^{n_j} \psi_{ij} \circ Y_{t-j}$$

Equation 4.A-3

where n_j is given by Equation 4.A-2.

When expressing the INARMA(p,q) process in terms of $\{Y_{t-j}\}_{j=1}^{\infty}$, it should be noted that because $\alpha \circ X + \beta \circ X \neq (\alpha + \beta) \circ X$, these coefficients cannot be added. Therefore, another summation over the number of Y_{t-j} terms has been used ($\sum_{i=1}^{n_j}$).

First, it is shown that the expression for n_j is correct when $j > p$. An example will motivate the general case. Consider an INARMA(2,1) process of:

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t + \beta_1 \circ Z_{t-1}$$

We are interested to find the number of Y_{t-3} terms in the IARR of this process. It can be seen that:

$$Z_{t-1} = Y_{t-1} - \alpha_1 \circ Y_{t-2} - \alpha_2 \circ Y_{t-3} - \beta_1 \circ Z_{t-2}$$

$$Z_{t-2} = Y_{t-2} - \alpha_1 \circ Y_{t-3} - \alpha_2 \circ Y_{t-4} - \beta_1 \circ Z_{t-3}$$

$$Z_{t-3} = Y_{t-3} - \alpha_1 \circ Y_{t-4} - \alpha_2 \circ Y_{t-5} - \beta_1 \circ Z_{t-4}$$

The INARMA(2,1) process can be written as:

$$\begin{aligned} Y_t &= \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t + \beta_1 \circ (Y_{t-1} - \alpha_1 \circ Y_{t-2} - \alpha_2 \circ Y_{t-3} - \beta_1 \circ Z_{t-2}) \\ &= \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t + \beta_1 \circ Y_{t-1} - \alpha_1\beta_1 \circ Y_{t-2} - \alpha_2\beta_1 \circ Y_{t-3} - \beta_1^2 \\ &\circ (Y_{t-2} - \alpha_1 \circ Y_{t-3} - \alpha_2 \circ Y_{t-4} - \beta_1 \circ Z_{t-3}) \\ &= \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t + \beta_1 \circ Y_{t-1} - \alpha_1\beta_1 \circ Y_{t-2} - \alpha_2\beta_1 \circ Y_{t-3} - \beta_1^2 \circ Y_{t-2} \\ &+ \alpha_1\beta_1^2 \circ Y_{t-3} + \alpha_2\beta_1^2 \circ Y_{t-4} + \beta_1^3 \circ (Y_{t-3} - \alpha_1 \circ Y_{t-4} - \alpha_2 \circ Y_{t-5} - \beta_1 \circ Z_{t-4}) \end{aligned}$$

It can be seen that Y_{t-3} terms come from Z_{t-1}, Z_{t-2}, and Z_{t-3} terms. It can be easily shown that Y_{t-4} terms come from Z_{t-2}, Z_{t-3}, and Z_{t-4} terms. Using the same argument, it can be shown that when $j > p$, Y_{t-j} terms come from $\{Z_{t-j+i}\}_{i=0}^{p}$ terms, with each of them producing one Y_{t-j}.

Now, we want to check if, for $j > p$, n_j is in fact $\sum_{i=1}^{q} n_{j-i}$ ($n_j = n_{j-1} + n_{j-2} + \cdots + n_{j-q}$). We know that n_j is the number of Y_{t-j} in IARR of an INARMA(p,q) process, so:

$n_{j-1} =$ the number of Y_{t-j+1}	$n_{j-2} =$ the number of Y_{t-j+2}	...	$n_{j-q} =$ the number of Y_{t-j+q}
which come from	which come from		which come from
↓	↓	...	↓
$\{Z_{t-j+1+i}\}_{i=0}^{p}$	$\{Z_{t-j+2+i}\}_{i=0}^{p}$	...	$\{Z_{t-j+q+i}\}_{i=0}^{p}$

The above terms can be written as follows:

Z_{t-j+1}	Z_{t-j+2}	$\dots$	Z_{t-j+q}
Z_{t-j+2}	Z_{t-j+3}	$\dots$	$Z_{t-j+q+1}$
Z_{t-j+3}	Z_{t-j+4}	$\dots$	$Z_{t-j+q+2}$
$\vdots$	$\vdots$		$\vdots$
$Z_{t-j+p+1}$	$Z_{t-j+p+2}$	$\dots$	$Z_{t-j+q+p}$

If we look at the elements that produce Y_{t-j} (which is the set $\{Z_{t-j+i}\}_{i=0}^{p} = \{Z_{t-j}, Z_{t-j+1}, \dots, Z_{t-j+p}\}$) we can see that:

1. the number of Z_{t-j} comes from $\{Z_{t-j+1}, Z_{t-j+2}, \dots, Z_{t-j+q}\}$ and as it can be seen in the above table, these are shown by the first rectangle.
2. the number of Z_{t-j+1} comes from $\{Z_{t-j+2}, Z_{t-j+3}, \dots, Z_{t-j+q+1}\}$ and as it can be seen in the above table, these are shown by the second rectangle.
3. the number of Z_{t-j+2} comes from $\{Z_{t-j+3}, Z_{t-j+4}, \dots, Z_{t-j+q+2}\}$ and as it can be seen in the above table, these are shown by the third rectangle.
4. the number of Z_{t-j+p} comes from $\{Z_{t-j+p+1}, Z_{t-j+p+2}, \dots, Z_{t-j+p+q}\}$ and as it can be seen in the above table, these are shown by the last rectangle.

Therefore, $n_j = n_{j-1} + n_{j-2} + \cdots + n_{j-q}$.

For the case of $0 < j \leq p$, the proof is the same except for we have one Y_{t-j} in the autoregressive (AR) part of Y_t itself. In the INARMA(2,1) example, if we are interested in the number of Y_{t-2} terms in the IARR of the process, it can be seen that there is one Y_{t-2} in the Y_t expression. In general, for $0 < j \leq p$, $n_j = \left(\sum_{i=1}^{q} n_{j-i}\right) + 1$.

In conclusion, the IARR for a general INARMA(p,q) process is given by Equation 4.A-1 with n_j determined by Equation 4.A-2.

Appendix 5.A The CLS Estimators of an INARMA(1,1) Model

In this appendix, the CLS estimates for the parameters of a PoINARMA(1,1) process are derived. As mentioned in section 5.7.2, the conditional least squares criterion is to minimize the following function with respect to the parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \lambda)'$ with other variables being taken as fixed:

$$Q_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})]^2$$

Equation 5.A-1

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \alpha} = \frac{\partial \{\sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})]^2\}}{\partial \alpha}$$
$$= -2 \sum_{t=1}^{n} Y_{t-1} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})] = 0$$

Equation 5.A-2

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \beta} = \frac{\partial \{\sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})]^2\}}{\partial \beta}$$
$$= -2 \sum_{t=1}^{n} Z_{t-1} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})] = 0$$

Equation 5.A-3

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \lambda} = \frac{\partial \{\sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})]^2\}}{\partial \lambda}$$
$$= -2 \sum_{t=1}^{n} [Y_t - (\alpha Y_{t-1} + \lambda + \beta Z_{t-1})] = 0$$

Equation 5.A-4

Solving the last equation results in the estimator for λ:

$$\hat{\lambda} = \frac{\sum_{t=1}^{n} Y_t - \hat{\alpha} \sum_{t=1}^{n} Y_{t-1} - \hat{\beta} \sum_{t=1}^{n} Z_{t-1}}{n}$$

Equation 5.A-5

The estimator for β can be found by substitution of $\hat{\lambda}$ in Equation 5.A-3:

$$\hat{\beta} = \frac{n\sum_{t=1}^{n} Y_t Z_{t-1} - n\hat{\alpha}\sum_{t=1}^{n} Y_{t-1}Z_{t-1} - \sum_{t=1}^{n} Y_t \sum_{t=1}^{n} Z_{t-1} + \hat{\alpha}\sum_{t=1}^{n} Y_{t-1}\sum_{t=1}^{n} Z_{t-1}}{n\sum_{t=1}^{n} Z_{t-1}^2 - \left(\sum_{t=1}^{n} Z_{t-1}\right)^2}$$

Equation 5.A-6

Finally, the estimator for α can be found by substitution of $\hat{\lambda}$ and $\hat{\beta}$ in Equation 5.A-2.

$$\sum_{t=1}^{n}\left[Y_tY_{t-1} - \left(\hat{\alpha}Y_{t-1}^2 + \hat{\lambda}Y_{t-1} + \hat{\beta}Y_{t-1}Z_{t-1}\right)\right]$$

$$= \sum_{t=1}^{n} Y_tY_{t-1} - \hat{\alpha}\sum_{t=1}^{n} Y_{t-1}^2 - \hat{\lambda}\sum_{t=1}^{n} Y_{t-1} - \hat{\beta}\sum_{t=1}^{n} Y_{t-1}Z_{t-1}$$

$$= \sum Y_tY_{t-1} - \hat{\alpha}\sum Y_{t-1}^2 - \frac{\sum Y_t - \hat{\alpha}\sum Y_{t-1} - \hat{\beta}\sum Z_{t-1}}{n}\sum Y_{t-1} - \hat{\beta}\sum Y_{t-1}Z_{t-1} = 0$$

Therefore,

$$n\sum Y_tY_{t-1} - n\hat{\alpha}\sum Y_{t-1}^2 - \sum Y_t\sum Y_{t-1} + \hat{\alpha}\left(\sum Y_{t-1}\right)^2$$
$$+\frac{n\sum Y_tZ_{t-1} - n\hat{\alpha}\sum Y_{t-1}Z_{t-1} - \sum Y_t\sum Z_{t-1} + \hat{\alpha}\sum Y_{t-1}\sum Z_{t-1}}{n\sum Z_{t-1}^2 - \left(\sum Z_{t-1}\right)^2}\left[\sum Y_{t-1}\sum Z_{t-1} - n\sum Y_{t-1}Z_{t-1}\right] = 0$$

So

$$n^2\hat{\alpha}\sum Y_{t-1}^2\sum Z_{t-1}^2 - n\hat{\alpha}\left(\sum Y_{t-1}\right)^2\sum Z_{t-1}^2 - n\hat{\alpha}\sum Y_{t-1}^2\left(\sum Z_{t-1}\right)^2$$
$$+\hat{\alpha}\left(\sum Y_{t-1}\right)^2\left(\sum Z_{t-1}\right)^2 + n\hat{\alpha}\sum Y_{t-1}\sum Z_{t-1}\sum Y_{t-1}Z_{t-1} - \hat{\alpha}\left(\sum Y_{t-1}\right)^2\left(\sum Z_{t-1}\right)^2$$
$$-n^2\hat{\alpha}\left(\sum Y_{t-1}Z_{t-1}\right)^2 + n\hat{\alpha}\sum Y_{t-1}\sum Z_{t-1}\sum Y_{t-1}Z_{t-1} =$$
$$n^2\sum Y_tY_{t-1}\sum Z_{t-1}^2 - n\sum Y_t\sum Y_{t-1}\sum Z_{t-1}^2 - n\sum Y_tY_{t-1}\left(\sum Z_{t-1}\right)^2$$
$$+\sum Y_t\sum Y_{t-1}\left(\sum Z_{t-1}\right)^2 + n\sum Y_{t-1}\sum Z_{t-1}\sum Y_tZ_{t-1}$$
$$-\sum Y_{t-1}\sum Y_t\left(\sum Z_{t-1}\right)^2 - n^2\sum Y_tZ_{t-1}\sum Y_{t-1}Z_{t-1} + n\sum Y_t\sum Z_{t-1}\sum Y_{t-1}Z_{t-1}$$

As a result, the estimator for α is given by:

$$\hat{\alpha} = \frac{\left\{\begin{matrix} n^2\sum Y_tY_{t-1}\sum Z_{t-1}^2 - n\sum Y_t\sum Y_{t-1}\sum Z_{t-1}^2 - n\sum Y_tY_{t-1}\left(\sum Z_{t-1}\right)^2 + n\sum Y_{t-1}\sum Z_{t-1}\sum Y_tZ_{t-1} \\ -n^2\sum Y_tZ_{t-1}\sum Y_{t-1}Z_{t-1} + n\sum Y_t\sum Z_{t-1}\sum Y_{t-1}Z_{t-1} \end{matrix}\right\}}{n^2\sum Y_{t-1}^2\sum Z_{t-1}^2 - n\left(\sum Y_{t-1}\right)^2\sum Z_{t-1}^2 - n\sum Y_{t-1}^2\left(\sum Z_{t-1}\right)^2 + 2n\sum Y_{t-1}\sum Z_{t-1}\sum Y_{t-1}Z_{t-1} - n^2\left(\sum Y_{t-1}Z_{t-1}\right)^2}$$

Equation 5.A-7

where all the summations are from 1 to n.

Appendix 5.B The Unconditional Variance of an INARMA(2,2) Model

The unconditional variance of an INARMA(2,2) process can be found from Equation 3-52 to be:

$$\mathrm{var}(Y_t) = \frac{\lambda}{1-\alpha_1^2-\alpha_2^2} \times \left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}\big(\alpha_1(1-\alpha_1)+\alpha_2(1-\alpha_2)\big)+\beta_1(1-\beta_1)+\beta_2(1-\beta_2)\right] + \frac{\lambda}{1-\alpha_1^2-\alpha_2^2}[1+\beta_1^2+\beta_2^2+2\alpha_1\beta_1+2\alpha_2\beta_2] + \frac{2\alpha_1\alpha_2\gamma_1+2\alpha_1\beta_2\gamma_1^{YZ}}{1-\alpha_1^2-\alpha_2^2}$$

Equation 5.B-1

From Equation 3-56, it can be seen that:

$$\gamma_1 = \alpha_1\gamma_0+\alpha_2\gamma_1+\beta_1\lambda+\beta_2\gamma_1^{YZ}$$

where γ_1^{YZ} is the cross-covariance at lag one given by Equation 3.C-1:

$$\gamma_1^{YZ} = (\alpha_1+\beta_1)\lambda$$

Therefore,

$$\gamma_1 = \frac{\alpha_1\gamma_0+\beta_1\lambda+\beta_2(\alpha_1+\beta_1)\lambda}{1-\alpha_2}$$

Equation 5.B-2

As a result, the variance of an INARMA(2,2) process is as follows:

$$\mathrm{var}(Y_t) = \frac{\lambda}{1-\alpha_1^2-\alpha_2^2} \times \left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}(\alpha_1-\alpha_1^2+\alpha_2-\alpha_2^2)+1+\beta_1+\beta_2+2\alpha_1\beta_1+2\alpha_2\beta_2\right] + \frac{2\alpha_1\alpha_2\frac{\alpha_1\gamma_0+\beta_1\lambda+\beta_2(\alpha_1+\beta_1)\lambda}{1-\alpha_2}+2\alpha_1\beta_2(\alpha_1+\beta_1)\lambda}{1-\alpha_1^2-\alpha_2^2}$$

$$\mathrm{var}(Y_t) = \frac{\lambda}{1-\alpha_1^2-\alpha_2^2} \times$$
$$\left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}(\alpha_1-\alpha_1^2+\alpha_2-\alpha_2^2)+1+\beta_1+\beta_2+2\alpha_1\beta_1+2\alpha_2\beta_2\right]$$
$$+\frac{2\alpha_1^2\alpha_2\gamma_0+2\alpha_1\alpha_2\beta_1\lambda+2\alpha_1\alpha_2\beta_2(\alpha_1+\beta_1)\lambda+2\alpha_1\beta_2(\alpha_1+\beta_1)(1-\alpha_2)\lambda}{(1-\alpha_1^2-\alpha_2^2)(1-\alpha_2)}$$

$$\mathrm{var}(Y_t) - \frac{2\alpha_1^2\alpha_2}{(1-\alpha_1^2-\alpha_2^2)(1-\alpha_2)}\mathrm{var}(Y_t) = \frac{\lambda}{1-\alpha_1^2-\alpha_2^2} \times$$
$$\left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}(\alpha_1-\alpha_1^2+\alpha_2-\alpha_2^2)+1+\beta_1+\beta_2+2\alpha_1\beta_1+2\alpha_2\beta_2\right]$$
$$+\frac{2\alpha_1\alpha_2\beta_1\lambda+2\alpha_1^2\beta_2\lambda+2\alpha_1\beta_1\beta_2\lambda}{(1-\alpha_1^2-\alpha_2^2)(1-\alpha_2)}$$

Finally, the unconditional variance of an INARMA(2,2) process is given by:

$$\mathrm{var}(Y_t) = \frac{\lambda}{(1-\alpha_1^2-\alpha_2^2)(1-\alpha_2)-2\alpha_1^2\alpha_2} \times$$
$$\left\{(1-\alpha_2)\left[\frac{1+\beta_1+\beta_2}{1-\alpha_1-\alpha_2}(\alpha_1-\alpha_1^2+\alpha_2-\alpha_2^2)+1+\beta_1+\beta_2+2\alpha_1\beta_1\right.\right.$$
$$\left.\left.+2\alpha_2\beta_2\right]+2\alpha_1\alpha_2\beta_1+2\alpha_1^2\beta_2+2\alpha_1\beta_1\beta_2\right\}$$

Equation 5.B-3

Appendix 6.A Lead Time Forecasting for an INAR(2) Model

In this appendix, we investigate forecasting of an INAR(2) process over a lead time $l \geq 1$. For the INAR(2) process of $Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + Z_t$, the cumulative Y over lead time l is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1} = (\alpha_1 \circ Y_t + \alpha_2 \circ Y_{t-1} + Z_{t+1})$$
$$+(\alpha_1 \circ Y_{t+1} + \alpha_2 \circ Y_t + Z_{t+2}) + \cdots + (\alpha_1 \circ Y_{t+l} + \alpha_2 \circ Y_{t+l-1} + Z_{t+l+1})$$

Equation 6.A-1

In order to find the conditional expectation of the aggregated process, Equation 6.A-1 should be expressed in terms of Y_t and Y_{t-1}.

Looking at the Equation 6.A-1, it can be seen that because Y_{t+1} is expressed in terms of Y_t and Y_{t-1}, there is no need for further substitution. For Y_{t+2}, there is one Y_{t+1} which should be expressed in terms of Y_t and Y_{t-1}. Therefore, the number of Y_t in the second element of the RHS of Equation 6.A-1 (obtained by repeated substitution of Y_{t+2}) is the number of Y_t in the first element plus one. This argument will be used in Appendix 6.C to find the number of $\{Y_{t-w+1}\}_{w=1}^{p}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. For other elements in the RHS of Equation 6.A-1, $\{Y_{t+j}\}_{j=3}^{l+1}$, the number of Y_t in each of them is equal to the sum of the number of Y_t in the previous two terms (Y_{t+j-1}, Y_{t+j-2}). The same argument applies for Y_{t-1}. Repeated substitution in Equation 6.A-1 yields:

$$\sum_{j=1}^{l+1} Y_{t+j} = (\alpha_1 \circ Y_t + \alpha_2 \circ Y_{t-1} + Z_{t+1})$$
$$+\left(\alpha_1^2 \circ Y_t + \alpha_1\alpha_2 \circ Y_{t-1} + \alpha_1 \circ Z_{t+1} + \alpha_2 \circ Y_t + Z_{t+2}\right)$$
$$+\left(\alpha_1^3 \circ Y_t + \alpha_1^2\alpha_2 \circ Y_{t-1} + \alpha_1^2 \circ Z_{t+1} + \alpha_1\alpha_2 \circ Y_t + \alpha_1 \circ Z_{t+2} + \alpha_1\alpha_2 \circ Y_t + \alpha_2^2 \circ Y_{t-1} + \alpha_2 \circ Z_{t+1} + Z_{t+3}\right) + \cdots$$

Equation 6.A-2

The above expression is not an infinite series but a finite series, where the remaining terms can be obtained by repeated substitution. The above equation can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Y_{t-1} + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^3} \psi_{ij}^3 \circ Z_{t+k_{ij}}$$

Equation 6.A-3

where n_j^1 is the number of Y_t terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in Equation 6.A-3 and ψ_{ij}^1 is the corresponding coefficient for each Y_t. n_j^2 is the number of Y_{t-1} terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in Equation 6.A-3 and ψ_{ij}^2 is the corresponding coefficient for each Y_{t-1}. n_j^3 is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in Equation 6.A-3 and ψ_{ij}^3 is the corresponding coefficient for each $Z_{t+k_{ij}}$. Each of these terms is explained below.

It can be seen from Equation 6.A-2 that because the process is an integer autoregressive of order two, the number of Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is equal to the sum of the number of Y_t in the two previous terms (Y_{t+j-1}, Y_{t+j-2}). Therefore $n_j^1 = n_{j-1}^1 + n_{j-2}^1$ for $j > 2$. As previously mentioned, when $j \leq 2$, the number of Y_t in Y_{t+j} is equal to the number of Y_t in the two previous terms plus one.

The corresponding coefficient for Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (say Y_{t+3}) is obtained from α_1 thinned the coefficient of Y_t in the previous term (in this case Y_{t+2}) and α_2 thinned the coefficient of Y_t in the next previous term (in this case Y_{t+1}). These coefficients are shown in Table 6.A-1.

Table 6.A-1 Coefficients of Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INAR(2) model

$j = 1, i = 1$	$\psi_{11}^1 = \alpha_1$
$j = 2, i = 1,2$	$\psi_{12}^1 = \alpha_1^2$ $\psi_{22}^1 = \alpha_2$
$\vdots$	$\vdots$
$j = l+1, i = 1, \ldots, n_{l+1}^1$	$\psi_{1(l+1)}^1 = \alpha_1^{l+1}$ $\psi_{2(l+1)}^1 = \alpha_1^{l-1}\alpha_2$ $\vdots$

The number of Y_{t-1} in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is also equal to the sum of the number of Y_t in the two previous terms (Y_{t+j-1}, Y_{t+j-2}). Therefore, $n_j^2 = n_{j-1}^2 + n_{j-2}^2$ for $j > 1$. But, for $j \leq 1$ this is equal to the number of Y_{t-1} in the two previous terms plus one.

The corresponding coefficient for Y_{t-1} in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (say Y_{t+3}) is obtained from α_1 thinned the coefficient of Y_{t-1} in the previous term (in this case Y_{t+2}) and α_2 thinned the coefficient of Y_{t-1} in the next previous term (in this case Y_{t+1}). These coefficients are shown in Table 6.A-2.

Table 6.A-2 Coefficients of Y_{t-1} in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INAR(2) model

$j = 1, i = 1$	$\psi_{11}^2 = \alpha_2$
$j = 2, i = 1$	$\psi_{12}^2 = \alpha_1\alpha_2$
$j = 3, i = 1,2$	$\psi_{13}^2 = \alpha_1^2\alpha_2$ $\psi_{23}^2 = \alpha_1\alpha_2$
$\vdots$	$\vdots$
$j = l + 1, i = 1, \dots, n_{l+1}^2$	$\psi_{1(l+1)}^2 = \alpha_1^l\alpha_2$ $\psi_{2(l+1)}^2 = \alpha_1^{l-1}\alpha_2$ $\vdots$

It can be seen from Equation 6.A-2 that because the process is an autoregressive process of order two, the number of $Z_{t+k_{ij}}$ increases in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. For the INAR(2) case, this number, shown by n_j^3, can be obtained from the number of $Z_{t+k_{ij}}$ in the two previous terms plus one because each of $\{Y_{t+j}\}_{j=1}^{l+1}$ has one Z_{t+j} as well. The corresponding coefficient for each $Z_{t+k_{ij}}$, shown by ψ_{ij}^3, is α_1 thinned the coefficients of $Z_{t+k_{ij}}$ in the previous term (Y_{t+j-1}) and α_2 thinned the coefficients of $Z_{t+k_{ij}}$ in the next previous term (Y_{t+j-2}). The coefficient for Z_{t+j} in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is one. $t + k_{ij}$ is the subscripts of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ which from Equation 6.A-2 it can be seen that k_{ij} is given by:

$$k_{ij} = \begin{cases} k_{i(j-2)} & \text{for } 1 \le i \le n_{j-2}^3 \\ k_{i(j-1)} & \text{for } n_{j-2}^3 < i \le n_{j-2}^3 + n_{j-1}^3 \\ j & \text{for } n_{j-2}^3 + n_{j-1}^3 < i \le n_j^3 \end{cases} \qquad \text{for } j = 1, \dots, l + 1$$

Equation 6.A-4

For example, for $j = 1$, the subscript of innovation term is $k_{11} = 1$, because as can be seen from Equation 6.A-1, there is only Z_{t+1} in the expression for Y_{t+1}. For $j = 2$, $Z_{t+k_{ij}}$s come from Y_{t+2} and Y_{t+1}, and it can be seen from Equation 6.A-2 that $k_{i2} = \begin{cases} 1 & i = 1 \\ 2 & i = 2 \end{cases}$. For $j > 2$, k_{ij} equals to all the subscripts included in the two previous $\{Y_{t+j}\}_{j=1}^{l+1}$ plus j. This means that k_{ij} is as shown in Equation 6.A-4.

The corresponding coefficient for each $Z_{t+k_{ij}}$ and the subscript of innovation terms, $t + k_{ij}$ are shown in Table 6.A-3.

Based on Equation 6.A-3, the conditional expected value of the aggregated process is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2\right) Y_{t-1} + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^3} \psi_{ij}^3\right) \lambda$$

Equation 6.A-5

Table 6.A-3 Coefficients of $Z_{t+k_{ij}}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INAR(2) model

$j = 1$ $i = 1, \dots, n_1^3$ where $n_1^3 = 1$	$\psi_{11}^3 = 1$	$k_{11} = 1$
$j = 2$ $i = 1, \dots, n_2^3$ where $n_2^3 = 2$	$\psi_{12}^3 = \alpha_1$ $\psi_{22}^3 = 1$	$k_{12} = 1$ $k_{22} = 2$
$j = 3$ $i = 1, \dots, n_3^3$ where $n_3^3 = 4$	$\psi_{13}^3 = \alpha_1^2$ $\psi_{23}^3 = \alpha_1$ $\psi_{33}^3 = \alpha_2$ $\psi_{43}^3 = 1$	$k_{13} = 1$ $k_{23} = 2$ $k_{33} = 1$ $k_{43} = 3$
$\vdots$	$\vdots$	$\vdots$
$j = l + 1$ $i = 1, \dots, n_{l+1}^3$	$\psi_{1(l+1)}^3 = \alpha_1^l$ $\psi_{2(l+1)}^3 = \alpha_1^{l-1}$ $\vdots$	$k_{1(l+1)} = 1$ $k_{2(l+1)} = 2$ $\vdots$

Appendix 6.B Lead Time Forecasting for an INARMA(1,2) Model

For the INARMA(1,2) process of $Y_t = \alpha \circ Y_{t-1} + Z_t + \beta_1 \circ Z_{t-1} + \beta_2 \circ Z_{t-2}$, the cumulative Y over lead time l is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1} = \left(\alpha \circ Y_t + Z_{t+1} + \beta_1 \circ Z_t + \beta_2 \circ Z_{t-1}\right)$$
$$+\left(\alpha \circ Y_{t+1} + Z_{t+2} + \beta_1 \circ Z_{t+1} + \beta_2 \circ Z_t\right) + \cdots +$$
$$+\left(\alpha \circ Y_{t+l} + Z_{t+l+1} + \beta_1 \circ Z_{t+l} + \beta_2 \circ Z_{t+l-1}\right)$$

Equation 6.B-1

In order to find the conditional expectation of the aggregated process, Equation 6.B-1 should be expressed in terms of Y_t.

$$\sum_{j=1}^{l+1} Y_{t+j} = (\alpha \circ Y_t + Z_{t+1} + \beta_1 \circ Z_t + \beta_2 \circ Z_{t-1})$$
$$+(\alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + \alpha\beta_1 \circ Z_t + \alpha\beta_2 \circ Z_{t-1} + Z_{t+2} + \beta_1 \circ Z_{t+1} + \beta_2 \circ Z_t)$$
$$+(\alpha^3 \circ Y_t + \alpha^2 \circ Z_{t+1} + \alpha^2\beta_1 \circ Z_t + \alpha^2\beta_2 \circ Z_{t-1} + \alpha \circ Z_{t+2} + \alpha\beta_1 \circ Z_{t+1}$$
$$+ \alpha\beta_2 \circ Z_t + Z_{t+3} + \beta_1 \circ Z_{t+2} + \beta_2 \circ Z_{t+1}) + \cdots$$

Equation 6.B-2

The above expression is not an infinite series but a finite series where the remaining terms can be obtained by repeated substitution. The above equation can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Z_{t+k_{ij}}$$

Equation 6.B-3

where n_j^1 is the number of Y_t terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ and ψ_{ij}^1 is the corresponding coefficient for each Y_t. n_j^2 is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$, ψ_{ij}^2 is the corresponding coefficient for each $Z_{t+k_{ij}}$. Each of these terms is explained below.

It can be seen from Equation 6.B-2 that because the process is an integer autoregressive of order one, each of $\{Y_{t+j}\}_{j=1}^{l+1}$ only has one Y_t and therefore $n_j^1 = 1$ (because the number of Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is equal to the number of Y_t in the previous term (Y_{t+j-1})). Therefore, the corresponding coefficient for Y_t in each of

$\{Y_{t+j}\}_{j=1}^{l+1}$ (say Y_{t+2}) is obtained from α thinned the coefficient of Y_t in the previous term (in this case Y_{t+1}). Therefore, $\psi_{ij}^1 = \alpha^j$. These coefficients are shown in Table 6.B-1.

Table 6.B-1 Coefficients of Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INARMA(1,2) model

$j = 1, i = 1$	$\psi_{11}^1 = \alpha$
$j = 2, i = 1$	$\psi_{12}^1 = \alpha^2$
$\vdots$	$\vdots$
$j = l + 1, i = 1$	$\psi_{1(l+1)}^1 = \alpha^{l+1}$

It can be seen from Equation 6.B-2 that because the process has an autoregressive component of order one and also a moving average component of order two, the number of $Z_{t+k_{ij}}$ increases in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. Each of $\{Y_{t+j}\}_{j=1}^{l+1}$ has three $Z_{t+k_{ij}}$ and also all the $Z_{t+k_{ij}}$ terms of the previous Y element (Y_{t+j-1}). Therefore, this number, shown by n_j^2, can be obtained from $n_{j-1}^2 + 3$. The same argument applies for the INARMA(p,q) case where the number of $Z_{t+k_{ij}}$ is equal to the number of them in the p previous terms plus $q + 1$ (see section 6.3.4). $t + k_{ij}$ is the subscripts of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ which from Equation 6.B-2 it can be seen that k_{ij} is given by:

$$k_{ij} = \begin{cases} k_{i(j-1)} & \text{for } i = 1, \ldots, n_{j-1}^2 \\ j, j-1, j-2 & \text{for } i = n_{j-1}^2 + 1, \ldots, n_j^2 \end{cases} \qquad \text{for } j = 1, \ldots, l+1$$

Equation 6.B-4

For example, for $j = 1$, the subscript of innovation terms is $k_{i1} = \begin{cases} 1 & i = 1 \\ 0 & i = 2 \\ -1 & i = 3 \end{cases}$, because as can be seen from Equation 6.B-1, there are three innovation terms in the expression for Y_{t+1} which are $\{Z_{t+1}, Z_t, Z_{t-1}\}$. For $j > 1$, the $Z_{t+k_{ij}}$s come from the innovation terms included in the previous $\{Y_{t+j}\}_{j=1}^{l+1}$ plus $\{j, j-1, j-2\}$. This means that k_{ij} is as shown in Equation 6.B-4.

The corresponding coefficient for each $Z_{t+k_{ij}}$ and the subscript of innovation terms, $t + k_{ij}$ are shown in Table 6.B-2.

Based on Equation 6.B-3, the conditional expected value of the aggregated process is:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j}\ |Y_t\right) = \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1}\psi_{ij}^1\right)Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2}\psi_{ij}^2\right)\lambda$$

Equation 6.B-5

Table 6.B-2 Coefficients of $Z_{t+k_{ij}}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ for an INARMA(1,2) model

$j=1$ $i=1,\dots,n_1^2$ where $n_1^2=3$	$\psi_{11}^2=1$ $\psi_{21}^2=\beta_1$ $\psi_{31}^2=\beta_2$		$k_{11}=1$ $k_{21}=0$ $k_{31}=-1$	
$j=2$ $i=1,\dots,n_2^2$ where $n_2^2=6$	$\psi_{12}^2=\alpha$ $\psi_{22}^2=\alpha\beta_1$ $\psi_{32}^2=\alpha\beta_2$	$\psi_{42}^2=1$ $\psi_{52}^2=\beta_1$ $\psi_{62}^2=\beta_2$	$k_{12}=1$ $k_{22}=0$ $k_{32}=-1$	$k_{42}=2$ $k_{52}=1$ $k_{62}=0$
$j=3$ $i=1,\dots,n_3^2$ where $n_3^2=9$	$\psi_{13}^2=\alpha^2$ $\psi_{23}^2=\alpha^2\beta_1$ $\psi_{33}^2=\alpha^2\beta_2$ $\psi_{43}^2=\alpha$ $\psi_{53}^2=\alpha\beta_1$	$\psi_{63}^2=\alpha\beta_2$ $\psi_{73}^2=1$ $\psi_{83}^2=\beta_1$ $\psi_{93}^2=\beta_2$	$k_{13}=1$ $k_{23}=0$ $k_{33}=-1$ $k_{43}=2$ $k_{53}=1$	$k_{63}=0$ $k_{73}=3$ $k_{83}=2$ $k_{93}=1$
$\vdots$	$\vdots$		$\vdots$	

Appendix 6.C Lead Time Forecasting for an INARMA(*p*,*q*) Model

In order to find the conditional mean of the over-lead-time-aggregated process, we need to express the aggregated INARMA(*p*,*q*) process in terms of the last p observations $(Y_{t-p+1}, Y_{t-p+2}, \dots, Y_{t-1}, Y_t)$. The aggregated process is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1}$$

Equation 6.C-1

Each of the $\{Y_{t+j}\}_{j=1}^{l+1}$ in the RHS of the above equation needs to be expressed in terms of $\{Y_{t-w+1}\}_{w=1}^{p}$ by repeated substitution of Y_{t+j} in Equation 3-50. Because the autoregressive order of the process is p, Y_{t+j} can be expressed in terms of p previous observations by $\alpha_1 \circ Y_{t+j-1} + \cdots + \alpha_p \circ Y_{t+j-p}$. Now, if $j \leq p-(w-1)$, as mentioned in Appendix 6.A, there is one $Y_{t-(w-1)}$ when we express the jth observation in the RHS of the Equation 6.C-1 (Y_{t+j}) without any need for repeated substitution. Repeated substitution of $(Y_{t+1}, \dots, Y_{t+p-(w-1)})$ by their p previous observations would result in obtaining more $Y_{t-(w-1)}$. Therefore, in total, the number of $Y_{t-(w-1)}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ when $j \leq p-(w-1)$ is equal to the number of $Y_{t-(w-1)}$ in its p previous observations plus one.

However, as explained in Appendix 6.A, when $j > p-(w-1)$, each Y_{t+j} from Equation 6.C-1 should be substituted by Equation 3-50 in order to reach $Y_{t-(w-1)}$, and the number of $Y_{t-(w-1)}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ would be equal to the number of $Y_{t-(w-1)}$ in its p previous observations.

For $j \leq p-(w-1)$, the corresponding coefficient of $Y_{t-(w-1)}$ in the jth observation in the RHS of the Equation 6.C-1 (Y_{t+j}) is $\alpha_{j+(w-1)}$ because:

$$Y_{t+j} = \alpha_1 \circ Y_{t+j-1} + \cdots + \alpha_{j+(w-1)} \circ Y_{t-(w-1)} + \cdots + \alpha_p \circ Y_{t+j-p} + Z_{t+j} + \sum_{i=1}^{q} \beta_i \circ Z_{t+j-i}$$

For other $Y_{t-(w-1)}$ the coefficient in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is α_i thinned the coefficient of $Y_{t-(w-1)}$ in the ith previous observation for $i = 1, \dots, p$.

For $j > p - (w - 1)$, again, the coefficient of $Y_{t-(w-1)}$ in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is α_i thinned the coefficient of $Y_{t-(w-1)}$ in the ith previous observation for $i = 1, \dots, p$ (the difference with the previous case is that we do not have $\alpha_{j+(w-1)}$).

Now we come back to Equation 6.C-1 to find the Z terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in the RHS of the equation when they expressed in terms of $\{Y_{t-w+1}\}_{w=1}^{p}$. As the process has a moving average component of order q, each $\{Y_{t+j}\}_{j=1}^{l+1}$ has $q + 1$ innovation terms $\{Z_{t+j}, Z_{t+j-1}, \dots, Z_{t+j-q}\}$. However, as mentioned in Appendix 6.B, by repeated substitution each $\{Y_{t+j}\}_{j=1}^{l+1}$ can be expressed in terms of p previous observations, each also with $q + 1$ innovation terms.

Therefore, the total number of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ is equal to the number of innovation terms in the p previous observations, plus $q + 1$. The corresponding coefficients for the $q + 1$ terms $\{Z_{t+j}, Z_{t+j-1}, \dots, Z_{t+j-q}\}$ are $\{1, \beta_1, \dots, \beta_q\}$, respectively. For the innovation terms that come from the p previous observations, coefficients would be α_k thinned the coefficient of $Z_{t+k_{ij}}$ in the kth previous observation for $k = 1, \dots, p$.

$t + k_{ij}$ denotes the subscript of Z for each i, j ($j = 1, \dots, l + 1$ and $i = 1, \dots, n_j^{p+1}$). As previously mentioned, each $\{Y_{t+j}\}_{j=1}^{l+1}$ has $q + 1$ innovation terms $\{Z_{t+j}, Z_{t+j-1}, \dots, Z_{t+j-q}\}$. Therefore, the subscripts for the last $q + 1$ innovation terms in each $\{Y_{t+j}\}_{j=1}^{l+1}$ are $\{j - q, j - 1, \dots, j\}$. This is shown in Table 6.C-1 by $i = n_{j-1}^{p+1} + 1, \dots, n_{j-1}^{p+1} + n_j^{p+1}$.

The other subscripts of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ simply are the subscripts of the innovation terms of p previous observations.

As a result, the aggregated process can be expressed as Equation 6-26 with the associated parameters as defined in Table 6.C-1.

Table 6.C-1 Parameters of the over-lead-time-aggregated INARMA(p,q) model

for $w = 1, \dots, p$	$n_j^w = \begin{cases} \left(\sum_{i=1}^{p} n_{j-i}^w\right) + 1 & j \le p - (w-1) \\ \sum_{i=1}^{p} n_{j-i}^w & j > p - (w-1) \end{cases}$	$\psi_{ij}^w = \begin{cases} \left[\begin{array}{ll} \alpha_p \psi_{i(j-p)}^w & i = 1, \dots, n_{j-p}^w \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^w & i = n_{j-2}^w + 1, \dots, n_{j-2}^w + n_{j-1}^w \\ \alpha_{j+(w-1)} & i = n_{j-1}^w + 1 \end{array}\right. & j \le p - (w-1) \\ \left[\begin{array}{ll} \alpha_p \psi_{i(j-p)}^w & i = 1, \dots, n_{j-p}^w \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^w & i = n_{j-2}^w + 1, \dots, n_{j-2}^w + n_{j-1}^w \end{array}\right. & j > p - (w-1) \end{cases}$
	$n_j^{p+1} = \left(\sum_{i=1}^{p} n_{j-i}^{p+1}\right) + (q+1)$	$\psi_{ij}^{p+1} = \begin{cases} \alpha_p \psi_{i(j-p)}^{p+1} & i = 1, \dots, n_{j-p}^{p+1} \\ \vdots & \vdots \\ \alpha_1 \psi_{i(j-1)}^{p+1} & i = n_{j-2}^{p+1} + 1, \dots, n_{j-2}^{p+1} + n_{j-1}^{p+1} \\ \beta_q, \dots, \beta_1, 1 & i = n_{j-1}^{p+1} + 1, \dots, n_{j-1}^{p+1} + n_j^{p+1} \end{cases}$ $k_{ij} = \begin{cases} \{k_{i(j-p)}\} & i = 1, \dots, n_{j-p}^{p+1} \\ \vdots & \vdots \\ \{k_{i(j-1)}\} & i = \sum_{z=2}^{p} n_{j-z}^{p+1} + 1, \dots, (\sum_{z=2}^{p} n_{j-z}^{p+1}) + n_{j-1}^{p+1} \\ j-q, \dots, j-1, j & i = \sum_{z=1}^{p} n_{j-z}^{p+1} + 1, \dots, n_j^{p+1} \end{cases}$

Appendix 8.A The MSE of YW and CLS Estimates for INAR(1), INMA(1) and INARMA(1,1) Processes

In this appendix, the two estimation methods used in this research are compared using MSE. Earlier in chapter 8, these methods have been compared in terms of their impact on forecast accuracy. The results for $n = 24, 36, 48, 96, 500$ are presented as follows.

Table 8.A-1 MSE of YW and CLS estimates of α for an INAR(1) process

Parameters	YW					CLS				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\lambda=0.5$	0.0184	0.0167	0.0144	0.0097	0.0031	0.0219	0.0186	0.0154	0.0100	0.0031
$\alpha=0.5,\lambda=0.5$	0.0742	0.0495	0.0359	0.0165	0.0030	0.0721	0.0481	0.0352	0.0161	0.0030
$\alpha=0.9,\lambda=0.5$	0.1460	0.0700	0.0435	0.0129	0.0009	0.1120	0.0519	0.0333	0.0104	0.0008
$\alpha=0.1,\lambda=1$	0.0196	0.0154	0.0138	0.0094	0.0026	0.0230	0.0167	0.0147	0.0097	0.0026
$\alpha=0.5,\lambda=1$	0.0725	0.0450	0.0333	0.0144	0.0028	0.0707	0.0440	0.0324	0.0140	0.0028
$\alpha=0.9,\lambda=1$	0.1546	0.0756	0.0456	0.0134	0.0008	0.1145	0.0582	0.0350	0.0109	0.0007
$\alpha=0.1,\lambda=3$	0.0192	0.0152	0.0136	0.0084	0.0026	0.0223	0.0167	0.0145	0.0087	0.0026
$\alpha=0.5,\lambda=3$	0.0679	0.0445	0.0319	0.0137	0.0021	0.0658	0.0430	0.0312	0.0135	0.0021
$\alpha=0.9,\lambda=3$	0.1468	0.0725	0.0442	0.0119	0.0008	0.1105	0.0555	0.0339	0.0095	0.0007
$\alpha=0.1,\lambda=5$	0.0192	0.0155	0.0129	0.0088	0.0027	0.0222	0.0170	0.0137	0.0091	0.0027
$\alpha=0.5,\lambda=5$	0.0709	0.0446	0.0320	0.0142	0.0022	0.0689	0.0429	0.0313	0.0139	0.0022
$\alpha=0.9,\lambda=5$	0.1455	0.0712	0.0437	0.0125	0.0008	0.1095	0.0550	0.0333	0.0101	0.0007

Table 8.A-2 Comparison of YW and CLS estimates of λ for an INAR(1) process

Parameters	YW					CLS				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\lambda=0.5$	0.0375	0.0243	0.0179	0.0100	0.0023	0.0380	0.0245	0.0181	0.0100	0.0023
$\alpha=0.5,\lambda=0.5$	0.1072	0.0712	0.0473	0.0224	0.0034	0.1030	0.0692	0.0464	0.0220	0.0034
$\alpha=0.9,\lambda=0.5$	3.9136	1.8446	1.1551	0.3370	0.0224	2.9900	1.3812	0.8930	0.2709	0.0206
$\alpha=0.1,\lambda=1$	0.0867	0.0608	0.0452	0.0272	0.0060	0.0898	0.0625	0.0463	0.0275	0.0060
$\alpha=0.5,\lambda=1$	0.3934	0.2350	0.1580	0.0713	0.0118	0.3852	0.2271	0.1545	0.0697	0.0118
$\alpha=0.9,\lambda=1$	16.5122	7.6779	4.7735	1.3540	0.0790	12.1860	5.8605	3.6707	1.1016	0.0721
$\alpha=0.1,\lambda=3$	0.3945	0.3001	0.2438	0.1413	0.0370	0.4275	0.3185	0.2528	0.1441	0.0372
$\alpha=0.5,\lambda=3$	2.8535	1.6850	1.2493	0.5244	0.0824	2.7658	1.6303	1.2121	0.5149	0.0822
$\alpha=0.9,\lambda=3$	132.6483	66.7297	40.2510	10.8786	0.6995	99.9406	51.2825	30.9385	8.7135	0.6386
$\alpha=0.1,\lambda=5$	0.8928	0.6958	0.5521	0.3577	0.0968	0.9817	0.7426	0.5772	0.3659	0.0972
$\alpha=0.5,\lambda=5$	7.4580	4.6569	3.4168	1.4914	0.2219	7.2026	4.4779	3.3367	1.4616	0.2204
$\alpha=0.9,\lambda=5$	368.6249	180.0181	111.2582	31.8673	2.0634	278.5108	139.5061	84.8648	25.7175	1.8826

Table 8.A-3 Comparison of YW and CLS estimates of β for an INMA(1) process

Parameters	YW					CLS				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.0516	0.0421	0.0349	0.0193	0.0039	0.0251	0.0216	0.0171	0.0097	0.0028
$\beta = 0.5, \lambda = 0.5$	0.1349	0.1008	0.0843	0.0493	0.0110	0.1147	0.0781	0.0721	0.0436	0.0211
$\beta = 0.9, \lambda = 0.5$	0.2161	0.1374	0.1024	0.0604	0.0121	0.2683	0.2008	0.1647	0.1131	0.0518
$\beta = 0.1, \lambda = 1$	0.0459	0.0364	0.0310	0.0166	0.0040	0.0236	0.0179	0.0152	0.0096	0.0029
$\beta = 0.5, \lambda = 1$	0.1233	0.0939	0.0822	0.0469	0.0106	0.1038	0.0805	0.0673	0.0417	0.0198
$\beta = 0.9, \lambda = 1$	0.2010	0.1485	0.1019	0.0548	0.0117	0.2752	0.2180	0.1726	0.1182	0.0517
$\beta = 0.1, \lambda = 3$	0.0468	0.0376	0.0297	0.0179	0.0038	0.0205	0.0172	0.0139	0.0096	0.0026
$\beta = 0.5, \lambda = 3$	0.1245	0.0960	0.0784	0.0474	0.0102	0.1110	0.0822	0.0696	0.0436	0.0207
$\beta = 0.9, \lambda = 3$	0.2141	0.1446	0.1156	0.0511	0.0113	0.3103	0.2391	0.2041	0.1321	0.0592
$\beta = 0.1, \lambda = 5$	0.0517	0.0400	0.0343	0.0154	0.0038	0.0214	0.0165	0.0151	0.0082	0.0027
$\beta = 0.5, \lambda = 5$	0.1222	0.0972	0.0829	0.0458	0.0095	0.1154	0.0902	0.0734	0.0427	0.0205
$\beta = 0.9, \lambda = 5$	0.2081	0.1388	0.0996	0.0563	0.0120	0.3434	0.2682	0.2226	0.1598	0.0695

Table 8.A-4 Comparison of YW and CLS estimates of λ for an INMA(1) process

Parameters	YW					CLS				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.0364	0.0240	0.0194	0.0103	0.0021	0.0370	0.0238	0.0189	0.0098	0.0020
$\beta = 0.5, \lambda = 0.5$	0.0599	0.0419	0.0280	0.0137	0.0025	0.0665	0.0456	0.0310	0.0160	0.0041
$\beta = 0.9, \lambda = 0.5$	0.0739	0.0461	0.0305	0.0153	0.0025	0.0781	0.0541	0.0376	0.0193	0.0050
$\beta = 0.1, \lambda = 1$	0.0830	0.0576	0.0467	0.0240	0.0057	0.0834	0.0562	0.0444	0.0220	0.0052
$\beta = 0.5, \lambda = 1$	0.1637	0.1042	0.0851	0.0393	0.0074	0.1831	0.1152	0.0948	0.0448	0.0124
$\beta = 0.9, \lambda = 1$	0.2442	0.1441	0.0993	0.0397	0.0065	0.2823	0.1737	0.1229	0.0540	0.0162
$\beta = 0.1, \lambda = 3$	0.3812	0.2835	0.2296	0.1407	0.0371	0.3530	0.2541	0.1971	0.1166	0.0314
$\beta = 0.5, \lambda = 3$	1.0334	0.7039	0.5304	0.2635	0.0507	1.1603	0.7754	0.5886	0.3137	0.1004
$\beta = 0.9, \lambda = 3$	1.5474	0.9868	0.7012	0.2655	0.0430	1.9331	1.3063	0.9528	0.4364	0.1303
$\beta = 0.1, \lambda = 5$	0.8853	0.6958	0.5688	0.3405	0.0880	0.7923	0.5717	0.4578	0.2725	0.0737
$\beta = 0.5, \lambda = 5$	2.6253	1.8145	1.4464	0.6367	0.1212	3.0335	2.1586	1.6302	0.7749	0.2682
$\beta = 0.9, \lambda = 5$	4.1058	2.4764	1.5930	0.7347	0.1130	5.8606	3.9471	2.7201	1.4349	0.4352

Table 8.A-5 Comparison of YW and CLS estimates of α for an INARMA(1,1) process

Parameters	YW					CLS				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.2421	0.2299	0.2029	0.1588	0.0474	0.0498	0.0420	0.0334	0.0225	0.0136
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.0590	0.0423	0.0326	0.0225	0.0085	0.0685	0.0576	0.0484	0.0329	0.0087
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.1426	0.1129	0.0904	0.0458	0.0065	0.1035	0.0743	0.0586	0.0308	0.0056
$\alpha=0.9,\beta=0.1,\lambda=0.5$	0.2587	0.1328	0.0806	0.0197	0.0011	0.1228	0.0654	0.0425	0.0129	0.0009
$\alpha=0.1,\beta=0.1,\lambda=1$	0.2440	0.2409	0.1968	0.1470	0.0492	0.0490	0.0380	0.0299	0.0246	0.0136
$\alpha=0.1,\beta=0.9,\lambda=1$	0.0666	0.0381	0.0336	0.0212	0.0081	0.0808	0.0643	0.0546	0.0337	0.0093
$\alpha=0.5,\beta=0.5,\lambda=1$	0.1444	0.1108	0.0829	0.0412	0.0062	0.0850	0.0621	0.0478	0.0257	0.0051
$\alpha=0.9,\beta=0.1,\lambda=1$	0.2593	0.1266	0.0744	0.0196	0.0011	0.1093	0.0571	0.0372	0.0111	0.0008
$\alpha=0.1,\beta=0.1,\lambda=5$	0.2295	0.2187	0.2025	0.1606	0.0428	0.0434	0.0334	0.0275	0.0200	0.0127
$\alpha=0.1,\beta=0.9,\lambda=5$	0.0606	0.0418	0.0317	0.0219	0.0082	0.1121	0.0965	0.0858	0.0631	0.0194
$\alpha=0.5,\beta=0.5,\lambda=5$	0.1373	0.1032	0.0843	0.0392	0.0061	0.0618	0.0441	0.0327	0.0183	0.0058
$\alpha=0.9,\beta=0.1,\lambda=5$	0.2611	0.1321	0.0758	0.0184	0.0012	0.1033	0.0554	0.0343	0.0101	0.0008

Table 8.A-6 Comparison of YW and CLS estimates of β for an INARMA(1,1) process

Parameters	YW					CLS				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.0470	0.0407	0.0353	0.0226	0.0109	0.0266	0.0205	0.0178	0.0118	0.0067
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.2646	0.1976	0.1597	0.0864	0.0158	0.4152	0.3694	0.3470	0.2715	0.1502
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.1520	0.1420	0.1405	0.1259	0.0495	0.1197	0.1128	0.1042	0.0948	0.0831
$\alpha=0.9,\beta=0.1,\lambda=0.5$	0.3237	0.2744	0.2342	0.1288	0.0189	0.0123	0.0100	0.0081	0.0063	0.0049
$\alpha=0.1,\beta=0.1,\lambda=1$	0.0492	0.0333	0.0305	0.0214	0.0106	0.0243	0.0181	0.0173	0.0119	0.0068
$\alpha=0.1,\beta=0.9,\lambda=1$	0.2876	0.2014	0.1632	0.0855	0.0163	0.4813	0.4255	0.3819	0.2880	0.1609
$\alpha=0.5,\beta=0.5,\lambda=1$	0.1488	0.1407	0.1386	0.1196	0.0460	0.1468	0.1280	0.1240	0.1146	0.0895
$\alpha=0.9,\beta=0.1,\lambda=1$	0.3443	0.2641	0.2224	0.1259	0.0164	0.0080	0.0075	0.0071	0.0063	0.0057
$\alpha=0.1,\beta=0.1,\lambda=5$	0.0455	0.0356	0.0296	0.0213	0.0118	0.0181	0.0152	0.0137	0.0108	0.0070
$\alpha=0.1,\beta=0.9,\lambda=5$	0.2685	0.2019	0.1575	0.0872	0.0149	0.6413	0.5912	0.5619	0.4657	0.2561
$\alpha=0.5,\beta=0.5,\lambda=5$	0.1492	0.1453	0.1421	0.1215	0.0483	0.1797	0.1752	0.1676	0.1621	0.1180
$\alpha=0.9,\beta=0.1,\lambda=5$	0.3464	0.2849	0.2332	0.1281	0.0184	0.0076	0.0075	0.0074	0.0074	0.0076

Table 8.A-7 Comparison of YW and CLS estimates of λ for an INARMA(1,1) process

Parameters	YW					CLS				
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.1007	0.0847	0.0731	0.0534	0.0124	0.0486	0.0315	0.0234	0.0136	0.0033
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.0993	0.0634	0.0466	0.0213	0.0049	0.0953	0.0643	0.0509	0.0289	0.0151
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.1929	0.1195	0.0785	0.0312	0.0045	0.1869	0.1295	0.0942	0.0485	0.0179
$\alpha=0.9,\beta=0.1,\lambda=0.5$	3.2137	1.4902	0.8934	0.2528	0.0200	2.6707	1.3440	0.8699	0.2635	0.0230
$\alpha=0.1,\beta=0.1,\lambda=1$	0.3580	0.3299	0.2642	0.1875	0.0484	0.1302	0.0933	0.0670	0.0396	0.0104
$\alpha=0.1,\beta=0.9,\lambda=1$	0.3386	0.2362	0.1530	0.0743	0.0174	0.3246	0.2378	0.1694	0.1023	0.0597
$\alpha=0.5,\beta=0.5,\lambda=1$	0.7878	0.4347	0.2709	0.1015	0.0160	0.7621	0.4629	0.3235	0.1661	0.0654
$\alpha=0.9,\beta=0.1,\lambda=1$	14.3119	6.4062	3.7534	1.0360	0.0795	12.9766	6.1504	3.9367	1.1341	0.0898
$\alpha=0.1,\beta=0.1,\lambda=5$	7.4699	6.9473	6.3751	4.8796	0.9735	1.6871	1.2283	0.9607	0.5643	0.1710
$\alpha=0.1,\beta=0.9,\lambda=5$	6.4788	4.2740	2.8742	1.5911	0.3502	5.7193	4.0864	2.9201	2.0613	1.3153
$\alpha=0.5,\beta=0.5,\lambda=5$	15.0720	9.3556	6.3524	2.3258	0.3487	15.1315	9.4722	6.8444	3.3042	1.3018
$\alpha=0.9,\beta=0.1,\lambda=5$	325.6348	154.4483	89.1639	25.0749	2.1841	300.4573	163.8322	101.8301	30.2243	2.5295

Appendix 8.B Impact of YW and CLS Estimates on Accuracy of Forecasts using MASE

The comparison of YW and CLS estimates for INAR(1), INMA(1) and INARMA(1,1) in terms of their impact on forecast accuracy using MASE is presented in this appendix.

Table 8.B-1 Forecast error comparison (YW and CLS) for INAR(1) series

Parameters	$\mathbf{MASE_{YW}/MASE_{CLS}}$					$\mathbf{MASE_{YW}/MASE_{CML}}$
	$n=24$	$n=36$	$n=48$	$n=96$	$n=500$	$n=24$
$\alpha=0.1,\lambda=0.5$	0.9888	0.9935	0.9975	0.9991	1.0000	0.9893
$\alpha=0.5,\lambda=0.5$	0.9983	1.0009	1.0001	1.0008	1.0002	1.0116
$\alpha=0.9,\lambda=0.5$	1.0837	1.0663	1.0483	1.0247	1.0038	1.1998
$\alpha=0.1,\lambda=1$	0.9858	0.9955	0.9973	0.9989	0.9999	0.9810
$\alpha=0.5,\lambda=1$	0.9942	0.9999	0.9993	1.0002	1.0001	1.0031
$\alpha=0.9,\lambda=1$	1.0593	1.0461	1.0364	1.0149	1.0012	1.1401
$\alpha=0.1,\lambda=3$	0.9916	0.9964	0.9979	0.9992	1.0000	0.9781
$\alpha=0.5,\lambda=3$	0.9955	1.0007	1.0015	0.9998	1.0000	0.9931
$\alpha=0.9,\lambda=3$	1.0462	1.0393	1.0265	1.0116	1.0005	1.1241
$\alpha=0.1,\lambda=5$	0.9914	0.9955	0.9979	0.9992	1.0000	0.9825
$\alpha=0.5,\lambda=5$	0.9956	0.9996	1.0001	1.0001	1.0000	1.0039
$\alpha=0.9,\lambda=5$	1.0500	1.0411	1.0292	1.0120	1.0006	-

Table 8.B-2 Forecast error comparison (YW and CLS) for INMA(1) series

Parameters	$\mathbf{MASE_{YW}/MASE_{CLS}}$				
	$\boldsymbol{n=24}$	$\boldsymbol{n=36}$	$\boldsymbol{n=48}$	$\boldsymbol{n=96}$	$\boldsymbol{n=500}$
$\beta=0.1,\lambda=0.5$	1.0029	1.0012	1.0013	1.0004	1.0001
$\beta=0.5,\lambda=0.5$	1.0043	1.0053	1.0030	1.0016	1.0001
$\beta=0.9,\lambda=0.5$	1.0060	1.0074	1.0050	1.0027	0.9975
$\beta=0.1,\lambda=1$	1.0011	1.0019	1.0011	1.0009	1.0005
$\beta=0.5,\lambda=1$	1.0092	1.0075	1.0062	1.0053	1.0041
$\beta=0.9,\lambda=1$	1.0099	1.0077	1.0083	1.0025	0.9971
$\beta=0.1,\lambda=3$	1.0004	1.0010	1.0015	1.0008	1.0003
$\beta=0.5,\lambda=3$	1.0025	1.0049	1.0032	1.0048	1.0020
$\beta=0.9,\lambda=3$	1.0001	0.9991	1.0006	1.0035	1.0007
$\beta=0.1,\lambda=5$	0.9993	1.0005	1.0006	1.0005	1.0001
$\beta=0.5,\lambda=5$	1.0004	1.0001	1.0003	1.0020	1.0019
$\beta=0.9,\lambda=5$	0.9933	0.9960	0.9972	0.9980	1.0018

Table 8.B-3 Comparison error comparison (YW and CLS) for INARMA(1,1) series

Parameters	$\mathbf{MASE_{YW}/MASE_{CLS}}$				
	$\boldsymbol{n=24}$	$\boldsymbol{n=36}$	$\boldsymbol{n=48}$	$\boldsymbol{n=96}$	$\boldsymbol{n=500}$
$\alpha=0.1, \beta=0.1, \lambda=0.5$	0.9956	0.9984	1.0069	1.0072	1.0023
$\alpha=0.1, \beta=0.9, \lambda=0.5$	1.0093	1.0250	1.0251	1.0212	1.0072
$\alpha=0.5, \beta=0.5, \lambda=0.5$	1.0233	1.0419	1.0386	1.0246	1.0049
$\alpha=0.9, \beta=0.1, \lambda=0.5$	1.1859	1.1255	1.0907	1.0368	1.0065
$\alpha=0.1, \beta=0.1, \lambda=1$	1.0289	1.0316	1.0336	1.0293	1.0121
$\alpha=0.1, \beta=0.9, \lambda=1$	1.0245	1.0364	1.0361	1.0283	1.0095
$\alpha=0.5, \beta=0.5, \lambda=1$	1.0617	1.0558	1.0478	1.0281	1.0066
$\alpha=0.9, \beta=0.1, \lambda=1$	1.1955	1.1250	1.0889	1.0295	1.0030
$\alpha=0.1, \beta=0.1, \lambda=5$	1.0504	1.0474	1.0474	1.0473	1.0109
$\alpha=0.1, \beta=0.9, \lambda=5$	1.0577	1.0643	1.0662	1.0625	1.0351
$\alpha=0.5, \beta=0.5, \lambda=5$	1.0845	1.0751	1.0749	1.0415	1.0121
$\alpha=0.9, \beta=0.1, \lambda=5$	1.3024	1.1553	1.1013	1.0309	1.0022

Appendix 8.C Croston-SBA Categorization for INAR(1), INMA(1) and INARMA(1,1)

The following tables show that the Croston-SBA categorization generally holds for INAR(1), INMA(1), and INARMA(1,1) processes although it originally developed for i.i.d. processes.

For INAR(1), when $\alpha = 0.1, \lambda = 0.5$, $\alpha = 0.5, \lambda = 0.5$, and $\alpha = 0.1, \lambda = 1$, SBA should outperform Croston based on the corresponding *p*-value.

Table 8.C-1 MSE of Croston and SBA with smoothing parameter 0.2 for INAR(1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.6782	0.6370	0.6204	0.6059	0.5888	0.6592	0.6248	0.6122	0.5986	0.5811
$\alpha = 0.5, \lambda = 0.5$	1.1889	1.1333	1.1098	1.0575	1.0550	1.1558	1.1027	1.0866	1.0333	1.0302
$\alpha = 0.9, \lambda = 0.5$	2.0661	2.1545	2.0588	2.0672	2.0163	2.3534	2.4621	2.3079	2.3119	2.2578
$\alpha = 0.1, \lambda = 1$	1.2918	1.2387	1.2419	1.2364	1.2016	1.2686	1.2244	1.2305	1.2236	1.1879
$\alpha = 0.5, \lambda = 1$	2.1306	2.0827	2.0617	2.0071	1.9784	2.1198	2.0755	2.0559	1.9977	1.9689
$\alpha = 0.9, \lambda = 1$	4.3286	4.1355	4.1565	4.0094	3.9999	5.4842	5.2874	5.2252	5.0125	4.9994
$\alpha = 0.1, \lambda = 3$	3.9531	3.8147	3.8414	3.7243	3.6340	4.0050	3.8476	3.8845	3.7627	3.6735
$\alpha = 0.5, \lambda = 3$	6.0937	5.8541	5.7937	5.7145	5.6107	6.3761	6.1316	6.0812	5.9983	5.8763
$\alpha = 0.9, \lambda = 3$	12.6744	12.5479	12.3444	12.0857	12.0343	21.6630	21.7462	21.4541	21.1452	21.0575
$\alpha = 0.1, \lambda = 5$	6.7050	6.3945	6.3156	6.1690	6.0777	6.9397	6.6147	6.5153	6.3704	6.2710
$\alpha = 0.5, \lambda = 5$	10.1928	9.6910	9.7351	9.3875	9.2467	11.1322	10.5665	10.6531	10.2523	10.0885
$\alpha = 0.9, \lambda = 5$	20.4023	20.8507	20.9265	20.1313	19.9123	48.8346	47.8702	46.9309	46.1660	44.8889

Table 8.C-2 MSE of Croston and SBA with smoothing parameter 0.5 for INAR(1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.7352	0.7163	0.7050	0.6894	0.6711	0.6719	0.6559	0.6474	0.6334	0.6152
$\alpha = 0.5, \lambda = 0.5$	1.1981	1.1722	1.1599	1.1137	1.1063	1.1030	1.0661	1.0646	1.0190	1.0096
$\alpha = 0.9, \lambda = 0.5$	1.3146	1.3137	1.2846	1.2601	1.2360	3.1301	3.1514	2.9995	2.9740	2.9217
$\alpha = 0.1, \lambda = 1$	1.4453	1.4006	1.3999	1.4021	1.3642	1.3495	1.3189	1.3243	1.3199	1.2818
$\alpha = 0.5, \lambda = 1$	2.0781	2.0287	2.0136	1.9739	1.9389	2.1293	2.0815	2.0602	2.0147	1.9791
$\alpha = 0.9, \lambda = 1$	2.7022	2.5724	2.5546	2.4636	2.4352	9.9577	9.5920	9.5507	9.0077	9.0156
$\alpha = 0.1, \lambda = 3$	4.5453	4.4165	4.4184	4.2813	4.1936	4.8183	4.6422	4.6675	4.5205	4.4314
$\alpha = 0.5, \lambda = 3$	5.8520	5.6100	5.5535	5.4866	5.3744	7.7889	7.4908	7.4554	7.3408	7.1728
$\alpha = 0.9, \lambda = 3$	8.1248	7.7172	7.5610	7.3578	7.3587	68.5686	67.3196	66.2151	65.3240	64.8094
$\alpha = 0.1, \lambda = 5$	7.7797	7.4433	7.3601	7.1823	7.0591	9.0522	8.7136	8.5879	8.3941	8.2367
$\alpha = 0.5, \lambda = 5$	9.7481	9.2986	9.3367	9.0251	8.8548	15.7458	15.0617	15.1057	14.6590	14.3696
$\alpha = 0.9, \lambda = 5$	12.8141	12.8380	12.6929	12.3732	12.1335	186.8114	180.7881	178.0403	175.0332	170.3159

For INMA(1), when $\beta = 0.1, \lambda = 0.5$, $\beta = 0.5, \lambda = 0.5$, $\beta = 0.9, \lambda = 0.5$, and $\beta = 0.1, \lambda = 1$, SBA should outperform Croston based on the corresponding p-value.

Table 8.C-3 MSE of Croston and SBA with smoothing parameter 0.2 for INMA(1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.6682	0.6264	0.6169	0.5950	0.5826	0.6468	0.6148	0.6095	0.5876	0.5754
$\beta = 0.5, \lambda = 0.5$	0.9637	0.9103	0.8507	0.8495	0.8195	0.9285	0.8867	0.8324	0.8315	0.8030
$\beta = 0.9, \lambda = 0.5$	1.2077	1.1498	1.1031	1.0834	1.0584	1.1547	1.1164	1.0764	1.0562	1.0325
$\beta = 0.1, \lambda = 1$	1.3089	1.2502	1.2256	1.2002	1.1860	1.2915	1.2344	1.2120	1.1869	1.1729
$\beta = 0.5, \lambda = 1$	1.7781	1.6990	1.6832	1.6235	1.6365	1.7507	1.6803	1.6615	1.6031	1.6171
$\beta = 0.9, \lambda = 1$	2.2710	2.1981	2.1295	2.0714	2.0319	2.2540	2.1770	2.1101	2.0549	2.0131
$\beta = 0.1, \lambda = 3$	4.0167	3.8200	3.7213	3.6679	3.6364	4.0619	3.8592	3.7651	3.7018	3.6722
$\beta = 0.5, \lambda = 3$	5.1518	4.9393	4.9167	4.7959	4.6883	5.2691	5.0550	5.0298	4.9097	4.8008
$\beta = 0.9, \lambda = 3$	6.2893	6.1923	6.0164	5.8912	5.8008	6.5575	6.4307	6.2320	6.1184	6.0191
$\beta = 0.1, \lambda = 5$	6.5944	6.5473	6.2777	6.1716	6.0266	6.8133	6.7225	6.4675	6.3628	6.2156
$\beta = 0.5, \lambda = 5$	8.4587	8.1070	8.0842	7.9465	7.7757	8.9173	8.5772	8.5611	8.3625	0.5754
$\beta = 0.9, \lambda = 5$	10.6487	10.1183	10.1251	9.8462	9.6391	11.4393	10.8493	10.8578	10.5863	0.8030

Table 8.C-4 MSE of Croston and SBA with smoothing parameter 0.5 for INMA(1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.7150	0.7091	0.6930	0.6787	0.6620	0.6528	0.6465	0.6407	0.6232	0.6088
$\beta = 0.5, \lambda = 0.5$	1.0583	1.0341	0.9870	0.9845	0.9506	0.9453	0.9228	0.8816	0.8793	0.8504
$\beta = 0.9, \lambda = 0.5$	1.3103	1.2889	1.2476	1.2304	1.2033	1.1635	1.1542	1.1219	1.0987	1.0740
$\beta = 0.1, \lambda = 1$	1.4580	1.4187	1.3924	1.3648	1.3498	1.3746	1.3317	1.3083	1.2818	1.2684
$\beta = 0.5, \lambda = 1$	1.9305	1.8791	1.8557	1.7775	1.7988	1.8414	1.7931	1.7662	1.6996	1.7163
$\beta = 0.9, \lambda = 1$	2.3641	2.3151	2.2397	2.1726	2.1380	2.3223	2.2708	2.2024	2.1439	2.1031
$\beta = 0.1, \lambda = 3$	4.6389	4.4053	4.2831	4.2405	4.2051	4.8967	4.6412	4.5360	4.4624	4.4243
$\beta = 0.5, \lambda = 3$	5.5047	5.3033	5.2630	5.1602	5.0283	6.3117	6.1039	6.0345	5.9233	5.7804
$\beta = 0.9, \lambda = 3$	6.4306	6.2870	6.1180	5.9501	5.8768	8.0736	7.8664	7.6395	7.4564	7.3493
$\beta = 0.1, \lambda = 5$	7.6307	7.6316	7.3059	7.1697	7.0108	8.8930	8.7858	8.4906	8.3299	8.1573
$\beta = 0.5, \lambda = 5$	9.0566	8.6844	8.6661	8.4742	8.3162	11.9868	11.6011	11.5405	11.2015	11.0221
$\beta = 0.9, \lambda = 5$	10.8019	10.1784	10.2150	9.9558	9.7360	15.8961	15.1546	15.1366	14.7591	14.4960

Finally, for INARMA(1,1), when $\alpha = 0.1, \beta = 0.1, \lambda = 0.5$, $\alpha = 0.1, \beta = 0.9, \lambda = 0.5$, $\alpha = 0.5, \beta = 0.5, \lambda = 0.5$, and $\alpha = 0.1, \beta = 0.1, \lambda = 1$, SBA should outperform Croston and the opposite is true for the rest.

Table 8.C-5 MSE of Croston and SBA with smoothing parameter 0.2 for INARMA(1,1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.7572	0.7285	0.6975	0.6797	0.6672	0.7344	0.7133	0.6858	0.6691	0.6570
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.4658	1.3971	1.2976	1.2855	1.2758	1.4061	1.3529	1.2653	1.2548	1.2440
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	2.0666	1.9352	1.9751	1.8888	1.8780	2.0127	1.8915	1.9394	1.8472	1.8370
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	2.4911	2.4166	2.4541	2.3439	2.3694	2.9636	2.7476	2.7803	2.6402	2.6659
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.4875	1.4329	1.3832	1.3637	1.3459	1.4650	1.4128	1.3656	1.3466	1.3300
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	2.5633	2.6071	2.5802	2.4522	2.4170	2.5445	2.5750	2.5632	2.4316	2.3971
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	3.8122	3.5193	3.5507	3.4400	3.3913	3.8405	3.5307	3.5720	3.4446	3.4046
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	4.9377	4.9217	4.8844	4.9002	4.6510	6.4337	6.3152	6.2141	6.1640	5.8469
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	7.3304	7.0908	6.9961	6.8049	6.6737	7.6299	7.3524	7.2568	7.0638	6.9243
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	12.2581	11.8649	11.8106	11.4788	11.3795	13.3116	12.8960	12.7726	12.3983	12.2912
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	17.0179	16.6291	16.8486	16.3143	16.1399	19.2541	18.8518	18.9279	18.3672	18.1258
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	25.9712	23.8322	24.3298	24.2873	23.3176	57.2282	56.2972	55.9089	54.7003	53.6411

Table 8.C-6 MSE of Croston and SBA with smoothing parameter 0.5 for INARMA(1,1) series

Parameters	$\mathbf{MSE_{Croston}}$					$\mathbf{MSE_{SBA}}$				
	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.8278	0.8255	0.7996	0.7832	0.7660	0.7468	0.7457	0.7237	0.7103	0.6956
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.5852	1.5845	1.4591	1.4463	1.4452	1.4111	1.3976	1.3085	1.2966	1.2900
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	1.9839	1.8595	1.9022	1.8264	1.8119	1.8514	1.7625	1.8132	1.7249	1.7129
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.4987	1.4757	1.4697	1.4072	1.4056	3.7822	3.5755	3.5684	3.4579	3.4478
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.6598	1.6111	1.5495	1.5359	1.5132	1.5622	1.5098	1.4599	1.4443	1.4251
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	2.6260	2.7332	2.6587	2.5459	2.5066	2.6141	2.6703	2.6471	2.5198	2.4867
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	3.3841	3.1385	3.1717	3.0929	3.0331	3.7330	3.4326	3.4858	3.3582	3.3188
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	3.0666	2.9571	2.9202	2.8722	2.7725	12.0092	11.4371	11.2427	10.9966	10.6574
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	8.2959	7.9922	7.8299	7.6759	7.5434	10.0345	9.6419	9.4752	9.2596	9.0982
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	12.2479	11.7657	11.7342	11.4066	11.3330	18.7567	18.1540	17.8894	17.4457	17.2565
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	14.8220	14.5978	14.7661	14.3685	14.1216	29.0148	28.5701	28.2932	27.7230	27.1435
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	15.6201	14.5525	14.5893	14.4396	13.9223	220.0655	216.2880	212.1152	207.3633	205.4631

Appendix 8.D Comparing the Accuracy of INARMA Forecasts for all points in time and issue points

In this appendix, the forecast accuracy of INARMA methods when all points in time are considered is compared to the case of issue points.

Table 8.D-1 Forecast accuracy for all points in time and issue points for INAR(1) series (known order)

Parameters	$n=24$						$n=36$						$n=48$						$n=96$					
	All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha=0.1,\lambda=0.5$	0.0026	0.6614	1.1418	-0.1034	0.8816	0.8592	-0.0089	0.6286	1.0555	-0.0840	0.7817	0.8115	0.0052	0.6047	1.0237	-0.0491	0.7214	0.7881	0.0045	0.5822	0.9822	-0.0247	0.6703	0.7545
$\alpha=0.5,\lambda=0.5$	-0.0026	0.9466	1.3290	-0.0241	1.2146	1.2289	-0.0089	0.8743	1.1807	-0.0316	1.0876	1.1100	0.0044	0.8532	1.1570	-0.0153	1.0373	1.0915	-0.0054	0.7834	1.0534	-0.0107	0.9473	0.9871
$\alpha=0.9,\lambda=0.5$	0.0017	1.2082	1.3901	-0.0028	1.2116	1.3893	0.0042	1.1489	1.2894	0.0013	1.1517	1.2895	-0.0043	1.0811	1.2176	-0.0058	1.0841	1.2172	-0.0033	1.0164	1.1486	-0.0047	1.0200	1.1492
$\alpha=0.1,\lambda=1$	-0.0149	1.2880	0.9376	-0.0962	1.4564	0.9781	-0.0059	1.2072	0.8496	-0.0582	1.2943	0.8580	0.0052	1.1967	0.8382	-0.0333	1.2759	0.8390	0.0029	1.1663	0.8057	-0.0132	1.2258	0.8090
$\alpha=0.5,\lambda=1$	-0.0195	1.8846	1.0837	-0.0428	2.0093	1.1169	-0.0018	1.7244	1.0202	-0.0126	1.8430	1.0457	-0.0045	1.6745	1.0059	-0.0104	1.7796	1.0307	0.0026	1.5905	0.9618	0.0002	1.6749	0.9852
$\alpha=0.9,\lambda=1$	-0.0121	2.5216	1.2401	-0.0125	2.5205	1.2400	0.0054	2.2655	1.1594	0.0053	2.2655	1.1594	-0.0146	2.1925	1.1262	-0.0147	2.1926	1.1262	-0.0105	2.0319	1.0686	-0.0107	2.0319	1.0686
$\alpha=0.1,\lambda=3$	0.0100	3.9224	0.8544	-0.0101	3.9851	0.8781	-0.0176	3.6703	0.8228	-0.0242	3.7034	0.8448	-0.0010	3.6641	0.8136	-0.0101	3.7006	0.8353	0.0085	3.4701	0.7846	0.0050	3.4883	0.8019
$\alpha=0.5,\lambda=3$	-0.0308	5.6926	1.0126	-0.0325	5.6986	1.0139	-0.0035	5.1557	0.9491	-0.0057	5.1609	0.9514	-0.0078	4.9742	0.9331	-0.0086	4.9801	0.9350	0.0143	4.7749	0.9160	0.0135	4.7793	0.9177
$\alpha=0.9,\lambda=3$	-0.0906	7.5862	1.1509	-0.0906	7.5862	1.1509	-0.0298	6.7494	1.1026	-0.0298	6.7494	1.1026	-0.0243	6.4376	1.0658	-0.0243	6.4376	1.0658	-0.0093	6.0230	1.0205	-0.0093	6.0230	1.0205
$\alpha=0.1,\lambda=5$	0.0371	6.6926	0.8565	0.0335	6.6920	0.8580	0.0209	6.1869	0.8143	0.0179	6.1929	0.8176	0.0282	6.0095	0.7973	0.0260	6.0176	0.8016	0.0139	5.7529	0.7742	0.0133	5.7586	0.7778
$\alpha=0.5,\lambda=5$	0.0123	9.3467	0.9840	0.0121	9.3479	0.9843	0.0082	8.6581	0.9653	0.0082	8.6597	0.9655	0.0116	8.3583	0.9392	0.0116	8.3583	0.9393	-0.0034	7.8699	0.9011	-0.0035	7.8699	0.9012
$\alpha=0.9,\lambda=5$	-0.0013	11.9986	1.1483	-0.0013	11.9986	1.1483	-0.0126	11.2051	1.0914	-0.0126	11.2051	1.0914	-0.0467	10.8985	1.0720	-0.0467	10.8985	1.0720	0.0185	10.1102	1.0272	0.0185	10.1102	1.0272

Table 8.D-2 Forecast accuracy for all points in time and issue points for INMA(1) series (known order)

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\beta = 0.1, \lambda = 0.5$	-0.0031	0.6295	1.1297	-0.0392	0.7823	0.8062	0.0027	0.6035	1.0473	0.0117	0.7134	0.7866	0.0114	0.5993	1.0493	0.0249	0.6832	0.7875	0.0007	0.5702	0.9736	0.0480	0.6570	0.7351
$\beta = 0.5, \lambda = 0.5$	0.0395	0.8793	1.3407	0.1612	1.1706	1.1189	0.0279	0.8552	1.2527	0.1975	1.1200	1.0541	0.0229	0.7997	1.1494	0.1969	1.0123	0.9901	0.0031	0.7885	1.0948	0.2116	0.9847	0.9574
$\beta = 0.9, \lambda = 0.5$	0.0767	1.1019	1.5472	0.2912	1.4609	1.3594	0.0671	1.0609	1.3632	0.2997	1.3175	1.2363	0.0628	1.0229	1.2606	0.3007	1.2400	1.1211	0.0445	0.9878	1.1797	0.3103	1.1874	1.0683
$\beta = 0.1, \lambda = 1$	0.0234	1.2748	0.9163	0.0114	1.3964	0.9335	0.0000	1.2038	0.8547	0.0050	1.2923	0.8731	0.0082	1.1724	0.8365	0.0223	1.2557	0.8414	0.0048	1.1313	0.7957	0.0391	1.2009	0.7956
$\beta = 0.5, \lambda = 1$	0.0347	1.7554	1.0666	0.1211	1.9307	1.1132	0.0440	1.6455	1.0247	0.1414	1.7795	1.0523	0.0310	1.6074	1.0005	0.1429	1.7315	1.0136	0.0204	1.5302	0.9648	0.1512	1.6235	0.9718
$\beta = 0.9, \lambda = 1$	0.1249	2.2869	1.2485	0.2374	2.4432	1.2942	0.1025	2.1650	1.1687	0.2334	2.2741	1.1922	0.1148	2.0762	1.1223	0.2549	2.1785	1.1295	0.1116	1.9944	1.0920	0.2551	2.0664	1.0959
$\beta = 0.1, \lambda = 3$	0.0452	3.9039	0.8547	0.0376	3.9375	0.8748	0.0402	3.6622	0.8158	0.0408	3.6950	0.8347	0.0150	3.5353	0.7960	0.0174	3.5579	0.8157	0.0062	3.4237	0.7767	0.0129	3.4386	0.7954
$\beta = 0.5, \lambda = 3$	0.1260	5.2415	0.9971	0.1379	5.2750	1.0039	0.0781	4.9565	0.9450	0.0896	4.9647	0.9540	0.0929	4.8942	0.9424	0.1089	4.9067	0.9486	0.0608	4.6821	0.9036	0.0760	4.6843	0.9094
$\beta = 0.9, \lambda = 3$	0.2970	6.7422	1.1268	0.3041	6.7525	1.1288	0.3209	6.5993	1.1023	0.3271	6.6089	1.1045	0.2796	6.2599	1.0587	0.2873	6.2574	1.0596	0.2718	6.0678	1.0364	0.2801	6.0598	1.0369
$\beta = 0.1, \lambda = 5$	0.0838	6.4549	0.8431	0.0830	6.4710	0.8465	0.0132	6.2194	0.8182	0.0129	6.2335	0.8213	0.0226	5.9292	0.7937	0.0239	5.9364	0.7972	0.0111	5.7446	0.7806	0.0138	5.7500	0.7836
$\beta = 0.5, \lambda = 5$	0.1407	8.7278	0.9676	0.1413	8.7273	0.9679	0.1984	8.3020	0.9417	0.1991	8.3003	0.9416	0.1738	8.0993	0.9245	0.1751	8.1031	0.9250	0.1269	7.8404	0.9063	0.1281	7.8395	0.9067
$\beta = 0.9, \lambda = 5$	0.4387	11.6802	1.1864	0.4387	11.6802	1.1865	0.4081	10.9042	1.0713	0.4083	10.9074	1.0714	0.4804	10.7677	1.0895	0.4807	10.7679	1.0895	0.4479	10.1863	1.0242	0.4481	10.1860	1.0243

Table 8.D-3 Forecast accuracy for all points in time and issue points for INARMA(1,1) series (known order)

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points			All points in time			Issue Points		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \beta = 0.1,$ $\lambda = 0.5$	-0.0007	0.7482	1.2163	-0.0700	1.0393	0.9354	-0.0066	0.7138	1.1186	-0.0382	0.9399	0.9074	0.0019	0.6777	1.0787	0.0039	0.8785	0.8779	-0.0047	0.6468	1.0076	0.0172	0.7936	0.8191
$\alpha = 0.1, \beta = 0.9,$ $\lambda = 0.5$	0.0401	1.2112	1.4173	0.1570	1.5961	1.3111	0.0336	1.1340	1.2974	0.1877	1.4754	1.1898	0.0267	1.0734	1.1958	0.1686	1.3287	1.0998	0.0163	1.0429	1.1248	0.2050	1.2677	1.0215
$\alpha = 0.5, \beta = 0.5,$ $\lambda = 0.5$	0.0485	1.5791	1.4864	0.1213	1.9469	1.3908	0.0209	1.4070	1.2666	0.0948	1.6876	1.2107	0.0445	1.3825	1.1924	0.1280	1.6513	1.1484	0.0261	1.2261	1.1137	0.1174	1.4511	1.0591
$\alpha = 0.9, \beta = 0.1,$ $\lambda = 0.5$	0.0920	1.4003	1.4228	0.0903	1.4066	1.4232	0.0532	1.2843	1.3142	0.0514	1.2898	1.3152	0.0487	1.2226	1.2608	0.0482	1.2273	1.2634	0.0324	1.1110	1.1489	0.0317	1.1134	1.1492
$\alpha = 0.1, \beta = 0.1,$ $\lambda = 1$	-0.0041	1.4908	0.9689	-0.0435	1.6925	1.0017	-0.0108	1.4146	0.9145	-0.0350	1.5629	0.9390	-0.0129	1.3358	0.8925	-0.0158	1.4591	0.9120	-0.0119	1.2833	0.8575	0.0018	1.3808	0.8725
$\alpha = 0.1, \beta = 0.9,$ $\lambda = 1$	0.0192	2.3265	1.1462	0.0618	2.5072	1.1789	-0.0185	2.2569	1.0930	0.0507	2.4059	1.1171	0.0368	2.2224	1.0713	0.1105	2.3482	1.0744	0.0149	2.0900	1.0210	0.1107	2.1709	1.0172
$\alpha = 0.5, \beta = 0.5,$ $\lambda = 1$	0.0296	3.1322	1.1914	0.0338	3.2529	1.2063	0.0250	2.7101	1.1033	0.0443	2.8128	1.1132	0.0409	2.6748	1.0798	0.0617	2.7683	1.0860	0.0220	2.4363	1.0245	0.0523	2.5181	1.0290
$\alpha = 0.9, \beta = 0.1,$ $\lambda = 1$	0.0665	2.8324	1.2617	0.0665	2.8324	1.2617	0.0494	2.5623	1.1748	0.0492	2.5623	1.1747	0.0640	2.4301	1.1219	0.0640	2.4301	1.1219	0.0413	2.2588	1.0786	0.0413	2.2587	1.0786
$\alpha = 0.1, \beta = 0.1,$ $\lambda = 5$	0.0279	7.4165	0.8824	0.0270	7.4271	0.8849	-0.0070	6.9501	0.8502	-0.0086	6.9563	0.8519	0.0011	6.7534	0.8307	-0.0008	6.7600	0.8323	0.0006	6.4253	0.8079	0.0011	6.4307	0.8101
$\alpha = 0.1, \beta = 0.9,$ $\lambda = 5$	0.0436	11.2147	1.0506	0.0436	11.2147	1.0506	0.0580	10.5085	0.9941	0.0580	10.5096	0.9942	0.0521	10.1101	0.9589	0.0521	10.1101	0.9589	0.0606	9.7366	0.9338	0.0606	9.7366	0.9338
$\alpha = 0.5, \beta = 0.5,$ $\lambda = 5$	0.1082	14.0754	1.0425	0.1082	14.0754	1.0425	0.1239	13.0227	1.0240	0.1239	13.0227	1.0240	0.1074	12.7060	0.9952	0.1074	12.7060	0.9952	0.1065	11.8259	0.9556	0.1065	11.8261	0.9556
$\alpha = 0.9, \beta = 0.1,$ $\lambda = 5$	0.1214	13.8720	1.1795	0.1214	13.8720	1.1795	0.1895	12.4636	1.0932	0.1895	12.4636	1.0932	0.1569	12.0389	1.0813	0.1569	12.0389	1.0813	0.0906	11.2782	1.0311	0.0906	11.2782	1.0311

Appendix 8.E Comparison of MASE of INARMA (known order) with Benchmarks

In this appendix, the degree of improvement by INARMA over benchmarks, using the MASE measure, is presented. The results are for the case where all points in time are taken into account.

Table 8.E-1 $MASE_{INARMA} / MASE_{Benchmark}$ for INARMA(0,0) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.8905	0.9259	0.9350	0.9269	0.9610	0.9658	0.9499	0.9841	0.9882	0.9670	1.0007	1.0046
$\lambda = 0.5$	0.9622	0.9837	0.9868	0.9795	0.9987	1.0010	0.9852	1.0033	1.0053	0.9866	1.0053	1.0074
$\lambda = 0.7$	0.9836	0.9954	0.9961	0.9873	0.9955	0.9961	0.9894	0.9969	0.9974	0.9881	0.9960	0.9966
$\lambda = 1$	0.9773	0.9944	0.9953	0.9698	0.9851	0.9858	0.9591	0.9750	0.9758	0.9448	0.9605	0.9613
$\lambda = 3$	0.9766	0.9853	0.9844	0.9669	0.9760	0.9753	0.9631	0.9722	0.9715	0.9538	0.9622	0.9614
$\lambda = 5$	0.9756	0.9753	0.9726	0.9649	0.9623	0.9593	0.9631	0.9620	0.9591	0.9544	0.9535	0.9508
$\lambda = 20$	0.9792	0.9146	0.8987	0.9652	0.9056	0.8906	0.9633	0.9011	0.8857	0.9564	0.8957	0.8803

Table 8.E-2 $MASE_{INARMA} / MASE_{Benchmark}$ for INMA(1) series (known order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9606	0.9810	0.9847	0.9775	0.9938	0.9956	0.9826	0.9981	0.9998	0.9867	1.0024	1.0041
$\beta = 0.5, \lambda = 0.5$	0.9459	0.9640	0.9665	0.9601	0.9751	0.9762	0.9660	0.9795	0.9805	0.9627	0.9772	0.9784
$\beta = 0.9, \lambda = 0.5$	0.9500	0.9717	0.9733	0.9453	0.9657	0.9679	0.9433	0.9617	0.9630	0.9243	0.9438	0.9453
$\beta = 0.1, \lambda = 1$	0.9757	0.9941	0.9953	0.9694	0.9900	0.9914	0.9603	0.9796	0.9808	0.9453	0.9657	0.9671
$\beta = 0.5, \lambda = 1$	0.9826	1.0020	1.0032	0.9828	1.0021	1.0034	0.9812	1.0010	1.0024	0.9796	1.0009	1.0026
$\beta = 0.9, \lambda = 1$	0.9955	1.0093	1.0098	0.9830	0.9974	0.9980	0.9786	0.9930	0.9935	0.9719	0.9864	0.9870
$\beta = 0.1, \lambda = 3$	0.9870	0.9934	0.9923	0.9740	0.9812	0.9802	0.9714	0.9796	0.9785	0.9675	0.9753	0.9743
$\beta = 0.5, \lambda = 3$	1.0073	1.0087	1.0060	0.9990	1.0010	0.9985	0.9977	0.9971	0.9943	0.9868	0.9878	0.9853
$\beta = 0.9, \lambda = 3$	1.0303	1.0191	1.0142	1.0296	1.0210	1.0164	1.0148	1.0087	1.0046	1.0089	1.0017	0.9973
$\beta = 0.1, \lambda = 5$	0.9883	0.9809	0.9771	0.9758	0.9756	0.9724	0.9690	0.9665	0.9632	0.9666	0.9625	0.9590
$\beta = 0.5, \lambda = 5$	1.0126	0.9957	0.9895	1.0091	0.9920	0.9855	0.9996	0.9825	0.9762	0.9906	0.9760	0.9700
$\beta = 0.9, \lambda = 5$	1.0493	1.0174	1.0084	1.0361	1.0119	1.0033	1.0252	1.0016	0.9933	1.0123	0.9875	0.9792

Table 8.E-3 $MASE_{INARMA} / MASE_{Benchmark}$ for INAR(1) series (known order)

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Cros 0.2	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**
$\alpha = 0.1, \lambda = 0.5$	0.9787	0.9505	0.9965	1.0121	0.3777	0.4554	0.9825	0.9381	0.9987	1.0027	1.0003	1.0329	0.9778	0.9312	0.9933	0.9957	0.9950	1.0273	0.9805	0.9342	0.9957	0.9979	0.9974	1.0279
$\alpha = 0.5, \lambda = 0.5$	0.8685	0.8610	0.8837	0.9055	0.8851	0.8811	0.8619	0.8470	0.8772	0.8956	0.8784	0.8728	0.8595	0.8365	0.8745	0.8857	0.8755	0.8672	0.8435	0.8218	0.8598	0.8714	0.8610	0.8552
$\alpha = 0.9, \lambda = 0.5$	0.7325	0.9403	0.6868	0.5770	0.6771	0.4419	0.7060	0.9252	0.6606	0.5598	0.6510	0.4228	0.6921	0.8938	0.6521	0.5511	0.6429	0.4191	0.6694	0.8741	0.6350	0.5394	0.6265	0.4093
$\alpha = 0.1, \lambda = 1$	0.9966	0.9365	1.0161	0.9858	1.0175	1.0012	0.9812	0.9161	1.0012	0.9632	1.0025	0.9828	0.9710	0.9071	0.9895	0.9545	0.9907	0.9708	0.9570	0.8934	0.9773	0.9415	0.9787	0.9612
$\alpha = 0.5, \lambda = 1$	0.9349	0.9482	0.9451	0.9571	0.9450	0.8936	0.9028	0.9227	0.9123	0.9256	0.9121	0.8622	0.8942	0.9079	0.9059	0.9180	0.9060	0.8574	0.8805	0.8954	0.8928	0.9039	0.8929	0.8467
$\alpha = 0.9, \lambda = 1$	0.7581	0.9606	0.6670	0.4577	0.6496	0.3447	0.7313	0.9358	0.6460	0.4430	0.6295	0.3328	0.7165	0.9215	0.6402	0.4357	0.6239	0.3283	0.7004	0.9012	0.6266	0.4306	0.6111	0.3247
$\alpha = 0.1, \lambda = 3$	0.9924	0.9297	0.9980	0.9202	0.9966	0.8996	0.9817	0.9136	0.9895	0.9117	0.9886	0.8982	0.9775	0.9152	0.9845	0.9078	0.9834	0.8899	0.9657	0.9028	0.9738	0.8988	0.9727	0.8794
$\alpha = 0.5, \lambda = 3$	0.9629	0.9835	0.9480	0.8547	0.9427	0.7377	0.9354	0.9586	0.9263	0.8362	0.9214	0.7231	0.9233	0.9463	0.9104	0.8199	0.9053	0.7095	0.9121	0.9333	0.9007	0.8129	0.8958	0.7042
$\alpha = 0.9, \lambda = 3$	0.7737	0.9664	0.5807	0.2813	0.5483	0.2114	0.7292	0.9333	0.5477	0.2692	0.5179	0.2021	0.7170	0.9195	0.5328	0.2631	0.5037	0.1975	0.7035	0.9023	0.5219	0.2572	0.4933	0.1930
$\alpha = 0.1, \lambda = 5$	1.0011	0.9328	0.9919	0.8695	0.9879	0.8169	0.9861	0.9142	0.9803	0.8533	0.9764	0.8033	0.9765	0.9051	0.9729	0.8484	0.9692	0.7991	0.9651	0.8974	0.9627	0.8415	0.9594	0.7954
$\alpha = 0.5, \lambda = 5$	0.9530	0.9735	0.9213	0.7596	0.9118	0.6289	0.9439	0.9638	0.9153	0.7520	0.9062	0.6239	0.9249	0.9464	0.8917	0.7361	0.8826	0.6100	0.9159	0.9355	0.8855	0.7255	0.8765	0.6026
$\alpha = 0.9, \lambda = 5$	0.7668	0.9666	0.4736	0.2105	0.4388	0.1575	0.7357	0.9335	0.4633	0.2066	0.4301	0.1548	0.7213	0.9263	0.4623	0.2057	0.4294	0.1541	0.7078	0.9048	0.4473	0.1990	0.4151	0.1491

Table 8.E-4 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ for INARMA(1,1) series (known order)

	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
Parameters	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9814	0.9497	1.0001	1.0118	1.0013	1.0305	0.9758	0.9275	0.9914	0.9938	0.9930	1.0190	0.9760	0.9256	0.9898	0.9893	0.9912	1.0136	0.9734	0.9184	0.9879	0.9850	0.9894	1.0135
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8861	0.8480	0.9101	0.9091	0.9105	0.9243	0.8839	0.8263	0.9048	0.8898	0.9065	0.9123	0.8869	0.8315	0.9066	0.8927	0.9082	0.9102	0.8753	0.8185	0.8963	0.8816	0.8980	0.9013
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.8580	0.8754	0.8728	0.9157	0.8726	0.8639	0.8286	0.8488	0.8451	0.8887	0.8463	0.8393	0.8177	0.8307	0.8341	0.8720	0.8353	0.8269	0.7894	0.8027	0.8086	0.8470	0.8101	0.8075
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.7329	0.9608	0.6755	0.5760	0.6648	0.4308	0.7079	0.9226	0.6649	0.5641	0.6555	0.4252	0.6793	0.8920	0.6412	0.5470	0.6324	0.4125	0.6602	0.8677	0.6275	0.5265	0.6189	0.3984
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9943	0.9404	1.0155	0.9913	1.0172	1.0070	0.9830	0.9236	1.0034	0.9746	1.0049	0.9951	0.9755	0.9196	0.9971	0.9694	0.9987	0.9868	0.9638	0.9072	0.9856	0.9579	0.9872	0.9777
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9421	0.9341	0.9531	0.9523	0.9530	0.9069	0.9172	0.9057	0.9325	0.9316	0.9332	0.9008	0.9145	0.9051	0.9283	0.9274	0.9287	0.8919	0.9127	0.9001	0.9282	0.9256	0.9288	0.8931
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8946	0.9514	0.8971	0.9190	0.8959	0.8149	0.8692	0.9220	0.8757	0.8987	0.8749	0.7999	0.8575	0.9107	0.8651	0.8892	0.8645	0.7934	0.8356	0.8852	0.8460	0.8682	0.8456	0.7767
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.7502	0.9519	0.6610	0.4423	0.6431	0.3323	0.7118	0.9250	0.6280	0.4284	0.6115	0.3217	0.6924	0.9060	0.6162	0.4199	0.6005	0.3153	0.6693	0.8805	0.6004	0.4114	0.5854	0.3093
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.0032	0.9442	0.9930	0.8617	0.9884	0.7962	0.9894	0.9338	0.9838	0.8610	0.9798	0.7976	0.9832	0.9325	0.9775	0.8558	0.9734	0.7890	0.9693	0.9142	0.9626	0.8407	0.9585	0.7795
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9532	0.9557	0.9246	0.7662	0.9161	0.6458	0.9347	0.9393	0.9065	0.7504	0.8981	0.6349	0.9226	0.9268	0.8982	0.7448	0.8902	0.6319	0.9176	0.9228	0.8923	0.7385	0.8843	0.6263
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9041	0.9718	0.8526	0.6689	0.8404	0.5331	0.8796	0.9427	0.8318	0.6528	0.8201	0.5215	0.8640	0.9250	0.8255	0.6533	0.8149	0.5235	0.8477	0.9035	0.8073	0.6340	0.7963	0.5079
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.7328	0.9408	0.4704	0.2077	0.4368	0.1559	0.7219	0.9226	0.4469	0.1978	0.4142	0.1482	0.7002	0.9074	0.4446	0.1971	0.4125	0.1478	0.6810	0.6836	0.2741	0.6155	0.2058	0.1446

Appendix 8.F Comparison of six-step ahead MSE of INARMA (known order) with Benchmarks

In this appendix the MSE of six-step ahead INARMA forecasts (using YW to estimate the parameters) is compared to the MSE of benchmark methods.

Table 8.F-1 Six-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ for INARMA(0,0) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\lambda=0.3$	0.8313	0.8826	0.8879	0.9378	0.9650	0.9675	0.9727	0.9891	0.9906	0.9789	0.9866	0.9873
$\lambda=0.5$	0.9364	0.9695	0.9723	0.9747	0.9907	0.9917	0.9782	0.9891	0.9896	0.9648	0.9744	0.9747
$\lambda=0.7$	0.9637	0.9882	0.9901	0.9751	0.9877	0.9880	0.9685	0.9787	0.9790	0.9581	0.9673	0.9674
$\lambda=1$	0.9735	0.9894	0.9898	0.9675	0.9782	0.9780	0.9659	0.9740	0.9736	0.9468	0.9552	0.9549
$\lambda=3$	0.9649	0.9548	0.9505	0.9478	0.9421	0.9382	0.9347	0.9295	0.9257	0.9206	0.9139	0.9100
$\lambda=5$	0.9621	0.9455	0.9384	0.9419	0.9200	0.9125	0.9373	0.9146	0.9070	0.9143	0.8932	0.8860
$\lambda=20$	0.9573	0.8218	0.7942	0.9389	0.8066	0.7794	0.9239	0.7887	0.7619	0.9142	0.7881	0.7619

Table 8.F-2 Six-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ for INMA(1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\beta=0.1,\lambda=0.5$	0.9173	0.9540	0.9572	0.9675	0.9856	0.9866	0.9669	0.9813	0.9821	0.9591	0.9705	0.9710
$\beta=0.5,\lambda=0.5$	0.9211	0.9695	0.9737	0.9503	0.9794	0.9815	0.9420	0.9647	0.9660	0.9239	0.9456	0.9470
$\beta=0.9,\lambda=0.5$	0.9292	0.9808	0.9853	0.9536	0.9849	0.9870	0.9432	0.9667	0.9681	0.9108	0.9346	0.9361
$\beta=0.1,\lambda=1$	0.9682	0.9902	0.9912	0.9622	0.9762	0.9764	0.9499	0.9626	0.9626	0.9368	0.9496	0.9496
$\beta=0.5,\lambda=1$	0.9668	0.9801	0.9797	0.9496	0.9675	0.9677	0.9342	0.9488	0.9487	0.9093	0.9231	0.9229
$\beta=0.9,\lambda=1$	0.9688	0.9939	0.9944	0.9256	0.9393	0.9388	0.9006	0.9190	0.9189	0.8832	0.8986	0.8982
$\beta=0.1,\lambda=3$	0.9577	0.9557	0.9519	0.9377	0.9248	0.9200	0.9216	0.9125	0.9081	0.9124	0.9060	0.9019
$\beta=0.5,\lambda=3$	0.9567	0.9403	0.9342	0.9184	0.9130	0.9079	0.9028	0.8885	0.8827	0.8804	0.8729	0.8679
$\beta=0.9,\lambda=3$	0.9476	0.9341	0.9269	0.9073	0.8973	0.8909	0.8862	0.8747	0.8683	0.8664	0.8512	0.8445
$\beta=0.1,\lambda=5$	0.9542	0.9369	0.9292	0.9301	0.9093	0.9016	0.9177	0.8968	0.8893	0.8990	0.8739	0.8661
$\beta=0.5,\lambda=5$	0.9579	0.9239	0.9130	0.9150	0.8823	0.8721	0.8984	0.8680	0.8580	0.8774	0.8470	0.8372
$\beta=0.9,\lambda=5$	0.9455	0.8851	0.8710	0.9027	0.8651	0.8529	0.8832	0.8417	0.8293	0.8586	0.8193	0.8075

Table 8.F-3 Six-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ with smoothing parameter 0.2 for INAR(1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\alpha=0.1,\lambda=0.5$	0.9290	0.9705	0.9745	0.9732	0.9921	0.9933	0.9714	0.9835	0.9841	0.9528	0.9655	0.9661
$\alpha=0.5,\lambda=0.5$	1.0424	1.0900	1.0939	0.9670	0.9962	0.9981	0.9315	0.9587	0.9605	0.8780	0.9073	0.9094
$\alpha=0.9,\lambda=0.5$	1.1764	1.1710	1.1607	1.0385	1.0363	1.0283	0.9826	0.9879	0.9810	0.8998	0.9066	0.9005
$\alpha=0.1,\lambda=1$	0.9826	1.0046	1.0055	0.9705	0.9836	0.9836	0.9520	0.9622	0.9619	0.9356	0.9485	0.9486
$\alpha=0.5,\lambda=1$	1.0263	1.0491	1.0492	0.9406	0.9667	0.9673	0.9170	0.9371	0.9371	0.8620	0.8837	0.8841
$\alpha=0.9,\lambda=1$	1.1586	1.0883	1.0665	1.0187	0.9664	0.9480	0.9825	0.9356	0.9185	0.8863	0.8386	0.8223
$\alpha=0.1,\lambda=3$	0.9748	0.9704	0.9664	0.9371	0.9315	0.9274	0.9255	0.9186	0.9143	0.9078	0.9008	0.8967
$\alpha=0.5,\lambda=3$	0.9864	0.9747	0.9676	0.9131	0.9114	0.9055	0.8737	0.8627	0.8564	0.8393	0.8310	0.8251
$\alpha=0.9,\lambda=3$	1.1322	0.8947	0.8454	1.0627	0.8301	0.7841	0.9961	0.7865	0.7450	0.8885	0.6957	0.6589
$\alpha=0.1,\lambda=5$	0.9460	0.9188	0.9108	0.9225	0.8997	0.8919	0.9174	0.8917	0.8837	0.9026	0.8786	0.8708
$\alpha=0.5,\lambda=5$	1.0105	0.9777	0.9643	0.9255	0.8918	0.8795	0.8862	0.8513	0.8394	0.8337	0.7990	0.7877
$\alpha=0.9,\lambda=5$	1.1459	0.7273	0.6712	1.0230	0.6764	0.6242	0.9755	0.6680	0.6178	0.8781	0.6001	0.5553

Table 8.F-4 Six-step ahead $\mathrm{MSE}_{\mathrm{INARMA}} / \mathrm{MSE}_{\mathrm{Benchmark}}$ with smoothing parameter 0.5 for INAR(1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$
$\alpha=0.1,\lambda=0.5$	0.8777	0.9762	0.9872	0.8689	0.9483	0.9536	0.8531	0.9259	0.9297	0.8261	0.9035	0.9095
$\alpha=0.5,\lambda=0.5$	0.9080	1.0391	1.0506	0.8214	0.9280	0.9327	0.7790	0.8829	0.8889	0.7437	0.8469	0.8534
$\alpha=0.9,\lambda=0.5$	1.1361	0.9462	0.7806	1.0338	0.8704	0.7239	0.9877	0.8391	0.6981	0.9112	0.7704	0.6399
$\alpha=0.1,\lambda=1$	0.8618	0.9320	0.9216	0.8344	0.8948	0.8832	0.8227	0.8788	0.8664	0.8092	0.8698	0.8589
$\alpha=0.5,\lambda=1$	0.8658	0.9195	0.8916	0.7825	0.8503	0.8275	0.7727	0.8259	0.8007	0.7184	0.7748	0.7540
$\alpha=0.9,\lambda=1$	1.1299	0.7420	0.5580	1.0106	0.6642	0.4991	0.9828	0.6461	0.4879	0.8961	0.5772	0.4317
$\alpha=0.1,\lambda=3$	0.8300	0.8018	0.7387	0.7874	0.7614	0.7013	0.7713	0.7441	0.6846	0.7603	0.7329	0.6749
$\alpha=0.5,\lambda=3$	0.8072	0.7399	0.6503	0.7424	0.6872	0.6053	0.7155	0.6538	0.5757	0.6844	0.6282	0.5531
$\alpha=0.9,\lambda=3$	1.1316	0.3808	0.2443	1.0441	0.3545	0.2283	1.0112	0.3477	0.2253	0.9042	0.3097	0.2011
$\alpha=0.1,\lambda=5$	0.7875	0.6921	0.6079	0.7670	0.6760	0.5899	0.7581	0.6655	0.5807	0.7516	0.6612	0.5766
$\alpha=0.5,\lambda=5$	0.8134	0.6490	0.5302	0.7560	0.6022	0.4922	0.7214	0.5764	0.4725	0.6793	0.5413	0.4441
$\alpha=0.9,\lambda=5$	1.1350	0.2570	0.1604	1.0247	0.2351	0.1454	0.9831	0.2347	0.1447	0.8907	0.2120	0.1311

Table 8.F-5 Six-step ahead $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.2 for INARMA(1,1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$	Croston $\alpha=0.2$	SBA $\alpha=0.2$	SBJ $A=0.2$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.9512	0.9912	0.9943	0.9824	1.0024	1.0038	0.9602	0.9795	0.9808	0.9486	0.9642	0.9650
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.9665	1.0085	1.0095	0.9345	0.9697	0.9722	0.9384	0.9654	0.9671	0.9016	0.9281	0.9299
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.9867	1.0228	1.0250	0.9133	0.9481	0.9506	0.8961	0.9309	0.9334	0.8562	0.8870	0.8890
$\alpha=0.9,\beta=0.1,\lambda=0.5$	1.0800	1.0416	1.0297	1.2237	1.2449	1.2379	1.1860	1.1924	1.1841	1.1816	1.1881	1.1794
$\alpha=0.1,\beta=0.1,\lambda=1$	1.0128	1.0355	1.0364	0.9989	1.0145	1.0147	0.9648	0.9791	0.9791	0.9254	0.9384	0.9384
$\alpha=0.1,\beta=0.9,\lambda=1$	0.9784	0.9922	0.9915	0.9274	0.9484	0.9485	0.9126	0.9302	0.9300	0.8686	0.8839	0.8835
$\alpha=0.5,\beta=0.5,\lambda=1$	0.9791	0.9998	0.9994	0.9318	0.9640	0.9648	0.8847	0.9072	0.9073	0.8367	0.8587	0.8588
$\alpha=0.9,\beta=0.1,\lambda=1$	1.2035	1.1125	1.0885	1.2078	1.1521	1.1302	1.2134	1.1574	1.1359	1.2180	1.1413	1.1183
$\alpha=0.1,\beta=0.1,\lambda=5$	0.9730	0.9560	0.9479	0.9283	0.9102	0.9024	0.9171	0.8995	0.8919	0.8933	0.8683	0.8601
$\alpha=0.1,\beta=0.9,\lambda=5$	0.9675	0.9382	0.9261	0.8961	0.8517	0.8392	0.8742	0.8333	0.8212	0.8486	0.8094	0.7974
$\alpha=0.5,\beta=0.5,\lambda=5$	0.9692	0.9024	0.8857	0.8892	0.8508	0.8372	0.8585	0.8164	0.8030	0.8123	0.7740	0.7614
$\alpha=0.9,\beta=0.1,\lambda=5$	1.3878	0.8918	0.8210	1.2591	0.8477	0.7841	1.3093	0.8560	0.7887	1.2348	0.8392	0.7758

Table 8.F-6 Six-step ahead $MSE_{INARMA} / MSE_{Benchmark}$ with smoothing parameter 0.5 for INARMA(1,1) series (known order)

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$
$\alpha=0.1,\beta=0.1,\lambda=0.5$	0.8927	0.9881	1.0203	0.8549	0.9433	0.9929	0.8271	0.9201	0.9682	0.8163	0.9023	0.9468
$\alpha=0.1,\beta=0.9,\lambda=0.5$	0.8253	0.9380	1.0095	0.7848	0.9007	0.9620	0.7870	0.8911	0.9425	0.7546	0.8594	0.9099
$\alpha=0.5,\beta=0.5,\lambda=0.5$	0.8752	0.9799	1.0064	0.7694	0.8795	0.9363	0.7577	0.8631	0.9183	0.7199	0.8149	0.8674
$\alpha=0.9,\beta=0.1,\lambda=0.5$	1.0323	0.8420	0.6669	1.2382	1.0821	0.8688	1.1947	1.0184	0.8128	1.1831	0.9947	0.7984
$\alpha=0.1,\beta=0.1,\lambda=1$	0.8833	0.9587	0.9878	0.8632	0.9266	0.9517	0.8292	0.8911	0.9180	0.7970	0.8552	0.8796
$\alpha=0.1,\beta=0.9,\lambda=1$	0.8156	0.8712	0.8932	0.7655	0.8308	0.8640	0.7558	0.8083	0.8436	0.7166	0.7668	0.7999
$\alpha=0.5,\beta=0.5,\lambda=1$	0.8055	0.8536	0.8870	0.7681	0.8332	0.8714	0.7397	0.7893	0.8140	0.6923	0.7404	0.7682
$\alpha=0.9,\beta=0.1,\lambda=1$	1.1949	0.7480	0.5465	1.2245	0.7934	0.5744	1.2326	0.8056	0.5853	1.2341	0.7862	0.5707
$\alpha=0.1,\beta=0.1,\lambda=5$	0.8034	0.7148	0.6574	0.7576	0.6714	0.6247	0.7533	0.6665	0.6185	0.7360	0.6413	0.5896
$\alpha=0.1,\beta=0.9,\lambda=5$	0.7717	0.6323	0.5574	0.7140	0.5658	0.4981	0.6934	0.5564	0.4883	0.6811	0.5392	0.4676
$\alpha=0.5,\beta=0.5,\lambda=5$	0.7894	0.5789	0.4843	0.7187	0.5541	0.4678	0.6910	0.5299	0.4482	0.6545	0.5015	0.4242
$\alpha=0.9,\beta=0.1,\lambda=5$	1.3669	0.3068	0.1871	1.2699	0.3015	0.1846	1.3210	0.2945	0.1798	1.2527	0.2946	0.1802

Appendix 8.G Comparison of all-INAR(1) and all-INARMA(1,1)

In this appendix, the forecast accuracy of all-INAR(1) and all-INARMA(1,1) is compared. The results are presented for the case of INAR(1), INMA(1) and INARMA(1,1) series. The corresponding results for an INARMA(0,0) series can be found from Table 8-32 and Table 8-43.

Table 8.G-1 Accuracy of forecasts by all-INAR(1) and all-INARMA(1,1) approaches for INAR(1) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \lambda = 0.5$	0.0026	0.6614	1.1418	-0.0008	0.6699	1.1275	-0.0089	0.6286	1.0555	0.0049	0.6447	1.0710	0.0052	0.6047	1.0237	0.0111	0.6389	1.0340	0.0045	0.5822	0.9822	0.0010	0.5857	0.9845
$\alpha = 0.5, \lambda = 0.5$	-0.0026	0.9466	1.3290	0.0130	1.0056	1.3278	-0.0089	0.8743	1.1807	0.0039	0.9436	1.2602	0.0044	0.8532	1.1570	0.0041	0.9131	1.1864	-0.0054	0.7834	1.0534	-0.0033	0.8186	1.0727
$\alpha = 0.9, \lambda = 0.5$	0.0017	1.2082	1.3901	0.0347	1.2681	1.4405	0.0042	1.1489	1.2894	0.0344	1.1681	1.3181	-0.0043	1.0811	1.2176	0.0066	1.0919	1.2150	-0.0033	1.0164	1.1486	0.0206	1.0235	1.1707
$\alpha = 0.1, \lambda = 1$	-0.0149	1.2880	0.9376	-0.0152	1.3755	0.9455	-0.0059	1.2072	0.8496	-0.0028	1.2778	0.8819	0.0052	1.1967	0.8382	-0.0066	1.1947	0.8427	0.0029	1.1663	0.8057	0.0045	1.1875	0.8235
$\alpha = 0.5, \lambda = 1$	-0.0195	1.8846	1.0837	-0.0092	1.9908	1.1188	-0.0018	1.7244	1.0202	0.0050	1.8491	1.0640	-0.0045	1.6745	1.0059	-0.0046	1.7735	1.0251	0.0026	1.5905	0.9618	0.0052	1.6469	0.9944
$\alpha = 0.9, \lambda = 1$	-0.0121	2.5216	1.2401	0.0544	2.4325	1.2304	0.0054	2.2655	1.1594	0.0378	2.3301	1.1748	-0.0146	2.1925	1.1262	0.0484	2.1853	1.1423	-0.0105	2.0319	1.0686	0.0312	2.0343	1.0747
$\alpha = 0.1, \lambda = 3$	0.0100	3.9224	0.8544	-0.0275	4.1292	0.8818	-0.0176	3.6703	0.8228	-0.0177	3.7457	0.8236	-0.0010	3.6641	0.8136	-0.0144	3.6442	0.8115	0.0085	3.4701	0.7846	-0.0222	3.4929	0.7828
$\alpha = 0.5, \lambda = 3$	-0.0308	5.6926	1.0126	0.0237	5.8020	1.0279	-0.0035	5.1557	0.9491	-0.0291	5.4831	0.9848	-0.0078	4.9742	0.9331	-0.0027	5.2166	0.9710	0.0143	4.7749	0.9160	0.0268	4.9748	0.9319
$\alpha = 0.9, \lambda = 3$	-0.0906	7.5862	1.1509	0.1097	7.3168	1.1422	-0.0298	6.7494	1.1026	0.1077	6.9143	1.1046	-0.0243	6.4376	1.0658	0.0627	6.5204	1.0742	-0.0093	6.0230	1.0205	0.0612	6.1440	1.0290
$\alpha = 0.1, \lambda = 5$	0.0371	6.6926	0.8565	0.0124	6.5975	0.8522	0.0209	6.1869	0.8143	0.0016	6.3370	0.8206	0.0282	6.0095	0.7973	-0.0078	6.2455	0.8197	0.0139	5.7529	0.7742	-0.0063	5.8121	0.7805
$\alpha = 0.5, \lambda = 5$	0.0123	9.3467	0.9840	0.0798	9.5203	1.0114	0.0082	8.6581	0.9653	0.0050	8.8544	0.9603	0.0116	8.3583	0.9392	0.0325	8.4735	0.9460	-0.0034	7.8699	0.9011	-0.0003	7.9638	0.9095
$\alpha = 0.9, \lambda = 5$	-0.0013	11.9986	1.1483	0.0773	12.2386	1.1385	-0.0126	11.2051	1.0914	0.1196	11.3739	1.0820	-0.0467	10.8985	1.0720	0.1021	10.9289	1.0679	0.0185	10.1102	1.0272	0.0956	10.1452	1.0232

Table 8.G-2 Accuracy of forecasts by all-INAR(1) and all-INARMA(1,1) approaches for INMA(1) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\beta = 0.1, \lambda = 0.5$	-0.0102	0.6916	1.1405	-0.0056	0.6500	1.1831	0.0016	0.6215	1.0690	-0.0038	0.6155	1.0589	-0.0084	0.5981	1.0149	-0.0071	0.6070	1.0025	0.00007	0.5687	0.9692	0.0063	0.5733	0.9809
$\beta = 0.5, \lambda = 0.5$	-0.0058	0.8285	1.2415	0.0205	0.8681	1.3428	0.0085	0.7828	1.1608	-0.0151	0.8047	1.1628	0.0033	0.7599	1.1155	0.0011	0.7815	1.1069	0.0038	0.6959	1.0381	-0.0038	0.7423	1.0637
$\beta = 0.9, \lambda = 0.5$	-0.0324	0.9442	1.2536	0.0430	1.0089	1.3941	-0.0052	0.8721	1.2436	0.0352	0.9383	1.2649	-0.0089	0.8149	1.1243	0.0122	0.9436	1.1917	-0.0063	0.7802	1.0677	0.0250	0.9208	1.1448
$\beta = 0.1, \lambda = 1$	-0.0005	1.3118	0.8950	-0.0202	1.3447	0.9356	-0.0078	1.2120	0.8617	-0.0154	1.2458	0.8487	0.0034	1.1882	0.8419	-0.0017	1.1956	0.8443	-0.0013	1.1519	0.8002	-0.0016	1.1525	0.8117
$\beta = 0.5, \lambda = 1$	-0.0271	1.6581	1.0424	0.0060	1.6743	1.0374	-0.0235	1.5574	0.9788	-0.0128	1.5905	0.9873	-0.0039	1.4772	0.9546	0.0030	1.5879	0.9894	-0.0047	1.4121	0.9168	-0.0077	1.4733	0.9413
$\beta = 0.9, \lambda = 1$	-0.0443	1.9014	1.1317	0.0145	1.9996	1.1350	-0.0203	1.7164	1.0562	0.0180	1.9324	1.0960	-0.0292	1.6602	1.0051	0.0165	1.8242	1.0545	-0.0127	1.5469	0.9562	0.0184	1.7825	1.0406
$\beta = 0.1, \lambda = 3$	0.0169	3.8869	0.8397	-0.0153	3.8953	0.8642	-0.0190	3.6392	0.8194	0.0022	3.7206	0.8211	-0.0208	3.4924	0.7922	0.0041	3.6216	0.7983	0.0136	3.4409	0.7838	-0.0215	3.5003	0.7873
$\beta = 0.5, \lambda = 3$	-0.0436	4.8555	0.9562	-0.0306	4.8945	0.9557	-0.0302	4.6371	0.9278	-0.0137	4.7351	0.9278	-0.0170	4.4562	0.8843	-0.0013	4.6193	0.9143	-0.0039	4.2607	0.8655	0.0061	4.3870	0.8783
$\beta = 0.9, \lambda = 3$	-0.1126	5.4157	1.0076	0.0763	5.8039	1.0545	-0.0603	5.1071	0.9687	0.0192	5.5915	1.0116	-0.0481	4.8830	0.9420	0.0118	5.2638	0.9726	-0.0213	4.7140	0.9137	0.0439	5.1879	0.9612
$\beta = 0.1, \lambda = 5$	-0.0330	6.3471	0.8369	0.0108	6.5867	0.8570	-0.0025	6.1832	0.8223	-0.0108	6.2327	0.8241	-0.0221	6.0512	0.7981	-0.0045	6.0357	0.7906	-0.0054	5.6957	0.7776	-0.0103	5.7236	0.7712
$\beta = 0.5, \lambda = 5$	-0.0071	8.0533	0.9474	-0.0335	8.3565	0.9718	-0.0317	7.5846	0.8934	0.0231	7.7461	0.9228	-0.0207	7.4066	0.8805	0.0136	7.6924	0.8954	-0.0166	6.8773	0.8510	0.0160	7.2509	0.8698
$\beta = 0.9, \lambda = 5$	-0.1464	9.3220	1.0338	0.0927	9.8826	1.0531	-0.0607	8.7192	0.9763	0.0643	8.8871	0.9798	-0.0626	8.1938	0.9370	0.0683	8.9059	0.9673	-0.0218	7.7859	0.9012	0.0524	8.4545	0.9443

Table 8.G-3 Accuracy of forecasts by all-INAR(1) and all-INARMA(1,1) approaches for INARMA(1,1) series

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)			All-INAR(1)			All-INARMA(1,1)		
	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE	ME	MSE	MASE
$\alpha = 0.1, \beta = 0.1,\ \lambda = 0.5$	-0.0078	0.7157	1.1853	-0.0007	0.7482	1.2163	0.0050	0.6776	1.0914	-0.0066	0.7138	1.1186	-0.0037	0.6789	1.0606	0.0019	0.6777	1.0787	0.0034	0.6319	1.0015	-0.0047	0.6468	1.0076
$\alpha = 0.1, \beta = 0.9,\ \lambda = 0.5$	-0.0336	1.1234	1.3333	0.0401	1.2112	1.4173	-0.0166	1.0234	1.1892	0.0336	1.1340	1.2974	-0.0108	0.9883	1.1387	0.0267	1.0734	1.1958	-0.0082	0.9030	1.0507	0.0163	1.0429	1.1248
$\alpha = 0.5, \beta = 0.5,\ \lambda = 0.5$	-0.0397	1.4167	1.3403	0.0485	1.5791	1.4864	-0.0108	1.3213	1.2244	0.0209	1.4070	1.2666	-0.0025	1.2331	1.1576	0.0445	1.3825	1.1924	-0.0109	1.1697	1.0806	0.0261	1.2261	1.1137
$\alpha = 0.9, \beta = 0.1,\ \lambda = 0.5$	0.0022	1.3138	1.3528	0.0920	1.4003	1.4228	-0.0171	1.2480	1.2641	0.0532	1.2843	1.3142	-0.0192	1.2192	1.2223	0.0487	1.2226	1.2608	0.0009	1.1111	1.1593	0.0324	1.1110	1.1489
$\alpha = 0.1, \beta = 0.1,\ \lambda = 1$	-0.0092	1.4535	0.9748	-0.0041	1.4908	0.9689	0.0050	1.3592	0.8997	-0.0108	1.4146	0.9145	-0.0022	1.3239	0.8900	-0.0129	1.3358	0.8925	0.0043	1.2502	0.8470	-0.0119	1.2833	0.8575
$\alpha = 0.1, \beta = 0.9,\ \lambda = 1$	-0.0421	2.2562	1.1278	0.0192	2.3265	1.1462	-0.0379	1.9904	1.0257	-0.0185	2.2569	1.0930	-0.0301	1.9530	1.0063	0.0368	2.2224	1.0713	-0.0113	1.8374	0.9588	0.0149	2.0900	1.0210
$\alpha = 0.5, \beta = 0.5,\ \lambda = 1$	-0.0791	2.9429	1.1710	0.0296	3.1322	1.1914	-0.0416	2.6021	1.0770	0.0250	2.7101	1.1033	-0.0334	2.4496	1.0386	0.0409	2.6748	1.0798	-0.0145	2.3081	0.9957	0.0220	2.4363	1.0245
$\alpha = 0.9, \beta = 0.1,\ \lambda = 1$	-0.0718	2.7062	1.2374	0.0665	2.8324	1.2617	-0.0625	2.4894	1.1512	0.0494	2.5623	1.1748	-0.0466	2.3962	1.1273	0.0640	2.4301	1.1219	-0.0156	2.2173	1.0745	0.0413	2.2588	1.0786
$\alpha = 0.1, \beta = 0.1,\ \lambda = 5$	-0.0116	7.2720	0.8837	0.0279	7.4165	0.8824	-0.0119	6.7888	0.8526	-0.0070	6.9501	0.8502	0.0285	6.6017	0.8402	0.0011	6.7534	0.8307	-0.0169	6.2483	0.8011	0.0006	6.4253	0.8079
$\alpha = 0.1, \beta = 0.9,\ \lambda = 5$	-0.1451	11.3162	1.0522	0.0436	11.2147	1.0506	-0.1048	10.1377	0.9675	0.0580	10.5085	0.9941	-0.0779	9.7491	0.9463	0.0521	10.1101	0.9589	-0.0438	9.1660	0.9021	0.0606	9.7366	0.9338
$\alpha = 0.5, \beta = 0.5,\ \lambda = 5$	-0.1845	14.2594	1.0796	0.1082	14.0754	1.0425	-0.1399	12.7866	1.0133	0.1239	13.0227	1.0240	-0.0804	12.2174	0.9767	0.1074	12.7060	0.9952	-0.0521	11.5310	0.9499	0.1065	11.8259	0.9556
$\alpha = 0.9, \beta = 0.1,\ \lambda = 5$	-0.3143	14.2380	1.1727	0.1214	13.8720	1.1795	-0.1687	12.8863	1.1125	0.1895	12.4636	1.0932	-0.0959	12.1338	1.0723	0.1569	12.0389	1.0813	-0.0613	11.0139	1.0149	0.0906	11.2782	1.0311

Appendix 8.H Comparison of MASE of INARMA (unknown order) with Benchmarks

In this appendix, the degree of improvement by using all-INAR(1) over benchmarks, in terms of the MASE is presented. The results are for the case where all points in time are taken into account. The results for INAR(1) series are the same as the case where the order is known (Table 8.E-3).

Table 8.H-1 $MASE_{INARMA} / MASE_{Benchmark}$ for INARMA(0,0) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.8989	0.9346	0.9438	0.9497	0.9846	0.9895	0.9533	0.9877	0.9918	0.9647	0.9983	1.0023
$\lambda = 0.5$	0.9777	0.9995	1.0027	0.9877	1.0071	1.0094	1.0007	1.0192	1.0212	1.0002	1.0192	1.0213
$\lambda = 0.7$	0.9992	1.0112	1.0119	0.9941	1.0024	1.0030	1.0107	1.0185	1.0189	0.9813	0.9892	0.9898
$\lambda = 1$	1.0102	1.0278	1.0287	0.9808	0.9963	0.9970	0.9691	0.9851	0.9860	0.9644	0.9804	0.9813
$\lambda = 3$	0.9967	1.0056	1.0047	0.9643	0.9734	0.9727	0.9577	0.9668	0.9660	0.9472	0.9555	0.9548
$\lambda = 5$	0.9688	0.9684	0.9658	0.9527	0.9502	0.9472	0.9745	0.9734	0.9704	0.9655	0.9646	0.9618
$\lambda = 20$	1.0024	0.9362	0.9199	0.9682	0.9083	0.8932	0.9815	0.9182	0.9025	0.9589	0.8980	0.8826

Table 8.H-2 $MASE_{INARMA} / MASE_{Benchmark}$ for INMA(1) series (unknown order)

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9698	0.9904	0.9941	0.9978	1.0144	1.0163	0.9504	0.9654	0.9670	0.9823	0.9978	0.9996
$\beta = 0.5, \lambda = 0.5$	0.8759	0.8927	0.8950	0.8897	0.9036	0.9045	0.9375	0.9506	0.9516	0.9129	0.9266	0.9277
$\beta = 0.9, \lambda = 0.5$	0.7697	0.7873	0.7886	0.8624	0.8810	0.8830	0.8413	0.8577	0.8589	0.8366	0.8542	0.8556
$\beta = 0.1, \lambda = 1$	0.9530	0.9710	0.9722	0.9773	0.9981	0.9995	0.9665	0.9859	0.9871	0.9507	0.9711	0.9725
$\beta = 0.5, \lambda = 1$	0.9603	0.9792	0.9804	0.9388	0.9572	0.9585	0.9362	0.9551	0.9564	0.9309	0.9511	0.9527
$\beta = 0.9, \lambda = 1$	0.9024	0.9149	0.9153	0.8884	0.9014	0.9019	0.8764	0.8893	0.8898	0.8510	0.8637	0.8642
$\beta = 0.1, \lambda = 3$	0.9696	0.9759	0.9749	0.9783	0.9856	0.9845	0.9668	0.9749	0.9738	0.9763	0.9842	0.9832
$\beta = 0.5, \lambda = 3$	0.9660	0.9673	0.9647	0.9809	0.9827	0.9803	0.9362	0.9357	0.9330	0.9452	0.9461	0.9437
$\beta = 0.9, \lambda = 3$	0.9213	0.9113	0.9069	0.9048	0.8973	0.8932	0.9029	0.8975	0.8938	0.8894	0.8831	0.8792
$\beta = 0.1, \lambda = 5$	0.9810	0.9737	0.9699	0.9807	0.9804	0.9773	0.9744	0.9719	0.9686	0.9629	0.9588	0.9553
$\beta = 0.5, \lambda = 5$	0.9914	0.9749	0.9688	0.9574	0.9411	0.9349	0.9520	0.9357	0.9298	0.9302	0.9164	0.9108
$\beta = 0.9, \lambda = 5$	0.9143	0.8865	0.8787	0.9442	0.9222	0.9143	0.8817	0.8614	0.8543	0.8907	0.8689	0.8616

Table 8.H-3 $\mathrm{MASE}_{\mathrm{INARMA}}$ / $\mathrm{MASE}_{\mathrm{Benchmark}}$ for INARMA(1,1) series (unknown order)

Parameters	$n = 24$						$n = 36$						$n = 48$						$n = 96$					
	Cros 0.2	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**	**Cros 0.2**	**Cros 0.5**	**SBA 0.2**	**SBA 0.5**	**SBJ 0.2**	**SBJ 0.5**
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9564	0.9255	0.9746	0.9860	0.9758	1.0042	0.9520	0.9050	0.9673	0.9696	0.9688	0.9943	0.9596	0.9101	0.9732	0.9727	0.9745	0.9966	0.9675	0.9129	0.9820	0.9791	0.9834	1.0073
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8336	0.7977	0.8562	0.8552	0.8565	0.8695	0.8102	0.7574	0.8293	0.8156	0.8309	0.8362	0.8445	0.7918	0.8633	0.8501	0.8648	0.8667	0.8176	0.7645	0.8372	0.8235	0.8389	0.8419
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.7737	0.7894	0.7870	0.8257	0.7868	0.7790	0.8010	0.8205	0.8170	0.8591	0.8181	0.8113	0.7938	0.8065	0.8098	0.8465	0.8109	0.8028	0.7659	0.7789	0.7845	0.8219	0.7860	0.7835
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.6969	0.9135	0.6423	0.5477	0.6321	0.4096	0.6809	0.8874	0.6396	0.5426	0.6305	0.4090	0.6586	0.8647	0.6216	0.5303	0.6131	0.3999	0.6662	0.8756	0.6332	0.5312	0.6245	0.4020
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0003	0.9461	1.0217	0.9973	1.0234	1.0131	0.9671	0.9087	0.9872	0.9589	0.9887	0.9790	0.9728	0.9171	0.9943	0.9667	0.9959	0.9841	0.9520	0.8961	0.9736	0.9462	0.9751	0.9657
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9270	0.9192	0.9378	0.9370	0.9377	0.8924	0.8607	0.8499	0.8751	0.8742	0.8757	0.8454	0.8591	0.8502	0.8720	0.8711	0.8724	0.8377	0.8571	0.8453	0.8716	0.8692	0.8722	0.8387
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8793	0.9351	0.8818	0.9033	0.8806	0.8009	0.8485	0.9001	0.8548	0.8773	0.8540	0.7808	0.8247	0.8759	0.8321	0.8553	0.8315	0.7631	0.8122	0.8603	0.8222	0.8438	0.8218	0.7548
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.7358	0.9336	0.6482	0.4337	0.6307	0.3259	0.6975	0.9064	0.6154	0.4198	0.5992	0.3153	0.6958	0.9104	0.6192	0.4219	0.6034	0.3168	0.6667	0.8771	0.5981	0.4099	0.5832	0.3081
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.0047	0.9456	0.9945	0.8630	0.9898	0.7973	0.9922	0.9364	0.9866	0.8635	0.9826	0.7998	0.9944	0.9432	0.9887	0.8656	0.9845	0.7980	0.9611	0.9065	0.9545	0.8336	0.9504	0.7730
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9546	0.9572	0.9260	0.7674	0.9175	0.6468	0.9096	0.9142	0.8823	0.7304	0.8741	0.6179	0.9105	0.9147	0.8864	0.7350	0.8785	0.6236	0.8864	0.8915	0.8620	0.7134	0.8543	0.6051
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9363	1.0064	0.8830	0.6927	0.8703	0.5520	0.8704	0.9329	0.8232	0.6460	0.8115	0.5161	0.8479	0.9078	0.8102	0.6411	0.7997	0.5138	0.8426	0.8981	0.8025	0.6302	0.7916	0.5048
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.7286	0.9354	0.4676	0.2065	0.4342	0.1550	0.7346	0.9389	0.4548	0.2012	0.4215	0.1508	0.6944	0.8999	0.4409	0.1955	0.4090	0.1465	0.6703	0.8707	0.4273	0.1897	0.3968	0.1423

Appendix 8.I Comparison of MASE of INARMA with Benchmarks for Lead Time Forecasts

In this appendix, the lead time forecasts of an all-INAR(1) method are compared to those of the benchmarks methods in terms of MASE. The results are presented for INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1) series and include both cases of $l = 3$ and $l = 6$.

Table 8.I-1 $\text{MASE}_{\text{INARMA}} / \text{MASE}_{\text{Benchmark}}$ of lead-time forecasts ($l = 3$) for INARMA(0,0) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.8692	0.9131	0.9178	0.9420	0.9748	0.9779	0.9692	0.9899	0.9917	0.9603	0.9773	0.9785
$\lambda = 0.5$	0.9484	0.9840	0.9874	0.9739	0.9993	1.0013	0.9742	0.9990	1.0010	0.9559	0.9827	0.9849
$\lambda = 0.7$	0.9755	1.0034	1.0054	0.9722	0.9968	0.9983	0.9622	0.9846	0.9860	0.9364	0.9581	0.9593
$\lambda = 1$	0.9824	0.9999	1.0000	0.9629	0.9819	0.9824	0.9516	0.9714	0.9719	0.9247	0.9481	0.9491
$\lambda = 3$	0.9624	0.9560	0.9512	0.9303	0.9263	0.9218	0.9121	0.9100	0.9058	0.8984	0.8963	0.8920
$\lambda = 5$	0.9567	0.9282	0.9184	0.9299	0.9056	0.8967	0.9112	0.8863	0.8775	0.8923	0.8710	0.8625
$\lambda = 20$	0.9552	0.7915	0.7606	0.9198	0.7657	0.7352	0.9044	0.7575	0.7286	0.8822	0.7413	0.7129

Table 8.I-2 $\text{MASE}_{\text{INARMA}} / \text{MASE}_{\text{Benchmark}}$ of lead-time forecasts ($l = 3$) for INMA(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9463	0.9858	0.9897	0.9682	0.9978	1.0003	0.9644	0.9914	0.9937	0.9506	0.9789	0.9812
$\beta = 0.5, \lambda = 0.5$	0.9748	1.0102	1.0134	0.9535	0.9838	0.9864	0.9443	0.9735	0.9759	0.9173	0.9462	0.9486
$\beta = 0.9, \lambda = 0.5$	0.9756	1.0186	1.0225	0.9587	0.9907	0.9933	0.9426	0.9727	0.9751	0.9066	0.9363	0.9388
$\beta = 0.1, \lambda = 1$	0.9854	1.0073	1.0082	0.9612	0.9833	0.9843	0.9455	0.9688	0.9699	0.9249	0.9493	0.9505
$\beta = 0.5, \lambda = 1$	0.9848	1.0053	1.0061	0.9577	0.9769	0.9774	0.9437	0.9646	0.9654	0.9120	0.9354	0.9364
$\beta = 0.9, \lambda = 1$	0.9974	1.0177	1.0183	0.9619	0.9824	0.9830	0.9447	0.9630	0.9632	0.9163	0.9376	0.9382
$\beta = 0.1, \lambda = 3$	0.9688	0.9579	0.9523	0.9391	0.9423	0.9386	0.9241	0.9209	0.9166	0.9010	0.9011	0.8972
$\beta = 0.5, \lambda = 3$	0.9849	0.9757	0.9701	0.9481	0.9412	0.9358	0.9369	0.9330	0.9281	0.9062	0.9027	0.8980
$\beta = 0.9, \lambda = 3$	0.9993	0.9771	0.9698	0.9592	0.9415	0.9344	0.9437	0.9346	0.9286	0.9103	0.9009	0.8949
$\beta = 0.1, \lambda = 5$	0.9686	0.9468	0.9378	0.9397	0.9124	0.9029	0.9151	0.8965	0.8882	0.9009	0.8794	0.8708
$\beta = 0.5, \lambda = 5$	0.9822	0.9528	0.9423	0.9451	0.9115	0.9014	0.9356	0.9060	0.8959	0.9077	0.8796	0.8701
$\beta = 0.9, \lambda = 5$	0.9935	0.9527	0.9404	0.9514	0.9110	0.8990	0.9358	0.9018	0.8906	0.9120	0.8763	0.8653

Table 8.I-3 $\mathrm{MASE}_{\mathrm{INARMA}} / \mathrm{MASE}_{\mathrm{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.2 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \lambda = 0.5$	0.9691	1.0016	1.0046	0.9605	0.9901	0.9927	0.9635	0.9896	0.9918	0.9461	0.9730	0.9753
$\alpha = 0.5, \lambda = 0.5$	0.9392	0.9627	0.9644	0.9229	0.9469	0.9489	0.9013	0.9267	0.9288	0.8690	0.8978	0.9002
$\alpha = 0.9, \lambda = 0.5$	0.8990	0.8433	0.8308	0.8421	0.8083	0.7986	0.8296	0.7975	0.7880	0.7843	0.7552	0.7468
$\alpha = 0.1, \lambda = 1$	0.9926	1.0157	1.0165	0.9528	0.9751	0.9760	0.9498	0.9743	0.9755	0.9222	0.9469	0.9481
$\alpha = 0.5, \lambda = 1$	0.9731	0.9895	0.9895	0.9413	0.9572	0.9572	0.9302	0.9459	0.9458	0.8970	0.9150	0.9152
$\alpha = 0.9, \lambda = 1$	0.9012	0.8115	0.7908	0.8522	0.7722	0.7543	0.8198	0.7565	0.7402	0.7881	0.7223	0.7063
$\alpha = 0.1, \lambda = 3$	0.9695	0.9666	0.9621	0.9391	0.9385	0.9343	0.9215	0.9226	0.9188	0.8995	0.8993	0.8953
$\alpha = 0.5, \lambda = 3$	0.9916	0.9694	0.9617	0.9533	0.9368	0.9297	0.9340	0.9227	0.9164	0.9070	0.8965	0.8905
$\alpha = 0.9, \lambda = 3$	0.8966	0.6667	0.6190	0.8333	0.6410	0.5952	0.8571	0.6316	0.6000	0.7975	0.6063	0.5748
$\alpha = 0.1, \lambda = 5$	0.9660	0.9418	0.9321	0.9221	0.9014	0.8931	0.9247	0.9028	0.8941	0.8955	0.8747	0.8664
$\alpha = 0.5, \lambda = 5$	0.9913	0.9520	0.9390	0.9533	0.9072	0.8942	0.9410	0.9081	0.8963	0.9097	0.8701	0.8580
$\alpha = 0.9, \lambda = 5$	0.8966	0.5778	0.5306	0.8571	0.5455	0.5000	0.8214	0.5349	0.5000	0.7857	0.5238	0.4889

Table 8.I-4 $\mathrm{MASE}_{\mathrm{INARMA}} / \mathrm{MASE}_{\mathrm{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.5 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \lambda = 0.5$	0.9137	0.9997	1.0076	0.8495	0.9461	0.9612	0.8498	0.9392	0.9513	0.8287	0.9197	0.9326
$\alpha = 0.5, \lambda = 0.5$	0.9021	0.9735	0.9739	0.8617	0.9443	0.9504	0.8390	0.9220	0.9271	0.8090	0.8950	0.9021
$\alpha = 0.9, \lambda = 0.5$	1.0239	0.6822	0.5505	1.0005	0.6886	0.5592	0.9720	0.6675	0.5411	0.9348	0.6491	0.5276
$\alpha = 0.1, \lambda = 1$	0.8763	0.9355	0.9231	0.8292	0.8940	0.8841	0.8287	0.8959	0.8863	0.8015	0.8664	0.8583
$\alpha = 0.5, \lambda = 1$	0.9283	0.9531	0.9171	0.8947	0.9200	0.8878	0.8806	0.9039	0.8682	0.8519	0.8800	0.8473
$\alpha = 0.9, \lambda = 1$	1.0284	0.5483	0.4237	1.0036	0.5384	0.4173	0.9687	0.5301	0.4097	0.9332	0.5023	0.3872
$\alpha = 0.1, \lambda = 3$	0.8164	0.7827	0.7180	0.7918	0.7595	0.6955	0.7806	0.7501	0.6889	0.7590	0.7305	0.6706
$\alpha = 0.5, \lambda = 3$	0.9383	0.8022	0.6979	0.8934	0.7717	0.6728	0.8767	0.7636	0.6670	0.8472	0.7448	0.6519
$\alpha = 0.9, \lambda = 3$	1.0400	0.3210	0.2407	1.0000	0.3165	0.2404	1.0000	0.3200	0.2400	0.9341	0.3033	0.2277
$\alpha = 0.1, \lambda = 5$	0.8048	0.7006	0.6078	0.7681	0.6762	0.5918	0.7750	0.6764	0.5888	0.7502	0.6559	0.5726
$\alpha = 0.5, \lambda = 5$	0.9238	0.7035	0.5795	0.8844	0.6729	0.5568	0.8755	0.6731	0.5566	0.8480	0.6448	0.5321
$\alpha = 0.9, \lambda = 5$	1.0400	0.2549	0.1926	1.0000	0.2449	0.1832	0.9583	0.2421	0.1811	0.9167	0.2366	0.1774

Table 8.I-5 $\mathrm{MASE}_{\mathrm{INARMA}} / \mathrm{MASE}_{\mathrm{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.2 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9599	0.9975	1.0010	0.9618	0.9911	0.9936	0.9600	0.9855	0.9876	0.9372	0.9661	0.9687
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.9729	1.0118	1.0154	0.9548	0.9863	0.9890	0.9307	0.9603	0.9629	0.9014	0.9314	0.9339
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9638	0.9822	0.9833	0.9093	0.9333	0.9351	0.8896	0.9166	0.9186	0.8623	0.8886	0.8906
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	0.8905	0.8351	0.8234	0.8429	0.7986	0.7881	0.8129	0.7814	0.7721	0.7725	0.7529	0.7452
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9896	1.0122	1.0133	0.9582	0.9806	0.9816	0.9460	0.9691	0.9701	0.9205	0.9448	0.9461
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0038	1.0170	1.0168	0.9593	0.9787	0.9790	0.9374	0.9554	0.9557	0.9109	0.9314	0.9320
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9807	0.9911	0.9900	0.9378	0.9517	0.9510	0.9083	0.9248	0.9246	0.8828	0.8981	0.8979
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	0.8947	0.7926	0.7720	0.8531	0.7740	0.7558	0.8164	0.7451	0.7283	0.7833	0.7197	0.7038
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9732	0.9490	0.9400	0.9484	0.9245	0.9152	0.9223	0.8989	0.8900	0.9046	0.8841	0.8756
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9920	0.9504	0.9376	0.9518	0.9100	0.8984	0.9261	0.8909	0.8798	0.9080	0.8755	0.8648
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9881	0.9263	0.9112	0.9412	0.8922	0.8781	0.9200	0.8737	0.8606	0.8924	0.8488	0.8359
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	0.9310	0.5625	0.5294	0.8621	0.5435	0.5102	0.8276	0.5333	0.5000	0.7857	0.5116	0.4783

Table 8.I-6 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 3)$ with smoothing parameter 0.5 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.8966	0.9951	1.0095	0.8618	0.9532	0.9666	0.8410	0.9323	0.9458	0.8194	0.9166	0.9328
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.8946	0.9929	1.0052	0.8521	0.9468	0.9584	0.8340	0.9284	0.9387	0.8031	0.8964	0.9079
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9370	0.9940	0.9853	0.8716	0.9462	0.9462	0.8430	0.9201	0.9190	0.8263	0.8974	0.8935
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.0438	0.6973	0.5664	0.9928	0.6744	0.5470	0.9687	0.6647	0.5385	0.9267	0.6488	0.5259
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.8902	0.9511	0.9374	0.8435	0.9077	0.8970	0.8267	0.8905	0.8809	0.8091	0.8737	0.8649
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9096	0.9439	0.9177	0.8608	0.9032	0.8812	0.8485	0.8884	0.8635	0.8222	0.8629	0.8402
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.9459	0.9530	0.9039	0.8983	0.9055	0.8584	0.8810	0.8928	0.8493	0.8492	0.8602	0.8193
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.0289	0.5346	0.4123	0.9976	0.5324	0.4109	0.9667	0.5169	0.3985	0.9323	0.5015	0.3862
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.8252	0.7119	0.6190	0.8066	0.6912	0.5989	0.7825	0.6771	0.5887	0.7676	0.6695	0.5832
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.8871	0.6957	0.5817	0.8536	0.6805	0.5728	0.8240	0.6598	0.5556	0.8071	0.6483	0.5468
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.9490	0.6764	0.5492	0.8956	0.6520	0.5311	0.8804	0.6434	0.5246	0.8565	0.6216	0.5054
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.0385	0.2523	0.1901	1.0000	0.2475	0.1852	1.0000	0.2424	0.1818	0.9167	0.2292	0.1732

Table 8.I-7 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ for INARMA(0,0) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\lambda = 0.3$	0.9262	0.9793	0.9846	0.8890	0.9311	0.9352	0.9549	0.9871	0.9898	0.9510	0.9776	0.9798
$\lambda = 0.5$	0.9864	1.0281	1.0313	0.9468	0.9776	0.9796	0.9496	0.9780	0.9796	0.9138	0.9451	0.9471
$\lambda = 0.7$	0.9598	1.0038	1.0066	0.9614	0.9907	0.9919	0.9425	0.9644	0.9648	0.8939	0.9233	0.9247
$\lambda = 1$	0.9710	1.0065	1.0077	0.9321	0.9571	0.9572	0.9145	0.9361	0.9356	0.8751	0.8965	0.8962
$\lambda = 3$	0.9355	0.9117	0.9022	0.8895	0.8703	0.8619	0.8704	0.8593	0.8518	0.8250	0.8159	0.8091
$\lambda = 5$	0.9362	0.8819	0.8666	0.8871	0.8368	0.8226	0.8589	0.8142	0.8004	0.8114	0.7758	0.7635
$\lambda = 20$	0.9167	0.7097	0.6667	0.8750	0.6774	0.6563	0.8696	0.6667	0.6250	0.7826	0.6000	0.5806

Table 8.I-8 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ for INMA(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\beta = 0.1, \lambda = 0.5$	0.9733	1.0270	1.0317	0.9555	0.9873	0.9895	0.9460	0.9688	0.9699	0.9049	0.9360	0.9382
$\beta = 0.5, \lambda = 0.5$	0.9523	1.0027	1.0072	0.9524	0.9850	0.9874	0.9383	0.9763	0.9792	0.8847	0.9200	0.9227
$\beta = 0.9, \lambda = 0.5$	0.9755	1.0382	1.0438	0.9576	0.9981	1.0012	0.9222	0.9570	0.9594	0.8849	0.9195	0.9219
$\beta = 0.1, \lambda = 1$	0.9687	0.9911	0.9911	0.9277	0.9523	0.9526	0.9085	0.9300	0.9298	0.8799	0.9028	0.9028
$\beta = 0.5, \lambda = 1$	0.9863	1.0066	1.0061	0.9355	0.9576	0.9575	0.9192	0.9416	0.9414	0.8598	0.8835	0.8837
$\beta = 0.9, \lambda = 1$	1.0103	1.0168	1.0145	0.9382	0.9550	0.9541	0.9144	0.9405	0.9405	0.8626	0.8844	0.8842
$\beta = 0.1, \lambda = 3$	0.9560	0.9298	0.9205	0.8952	0.8736	0.8655	0.8626	0.8468	0.8394	0.8220	0.8114	0.8048
$\beta = 0.5, \lambda = 3$	0.9711	0.9415	0.9315	0.9095	0.8958	0.8877	0.8734	0.8574	0.8500	0.8389	0.8249	0.8177
$\beta = 0.9, \lambda = 3$	0.9943	0.9640	0.9526	0.9231	0.9063	0.8971	0.8925	0.8704	0.8609	0.8399	0.8205	0.8119
$\beta = 0.1, \lambda = 5$	0.9580	0.9208	0.9069	0.8882	0.8398	0.8256	0.8529	0.8147	0.8021	0.8188	0.7821	0.7696
$\beta = 0.5, \lambda = 5$	0.9740	0.9160	0.8988	0.9034	0.8376	0.8217	0.8770	0.8353	0.8218	0.8360	0.7921	0.7788
$\beta = 0.9, \lambda = 5$	0.9902	0.9220	0.9034	0.9071	0.8420	0.8254	0.8766	0.8191	0.8032	0.8424	0.7897	0.7751

Table 8.I-9 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.2 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \lambda = 0.5$	0.9080	0.9615	0.9663	0.9518	0.9839	0.9863	0.9521	0.9779	0.9794	0.9064	0.9400	0.9424
$\alpha = 0.5, \lambda = 0.5$	0.9824	1.0279	1.0320	0.9527	0.9852	0.9877	0.9140	0.9488	0.9514	0.8656	0.9005	0.9032
$\alpha = 0.9, \lambda = 0.5$	1.0092	0.9926	0.9827	0.9498	0.9300	0.9211	0.9263	0.9104	0.9027	0.8765	0.8602	0.8525
$\alpha = 0.1, \lambda = 1$	0.9865	1.0081	1.0078	0.9450	0.9617	0.9612	0.9151	0.9374	0.9374	0.8687	0.8929	0.8933
$\alpha = 0.5, \lambda = 1$	1.0050	1.0159	1.0145	0.9551	0.9701	0.9692	0.9125	0.9373	0.9378	0.8690	0.8939	0.8942
$\alpha = 0.9, \lambda = 1$	1.0143	0.9115	0.8898	0.9557	0.8725	0.8532	0.9240	0.8675	0.8519	0.8679	0.8054	0.7894
$\alpha = 0.1, \lambda = 3$	0.9546	0.9386	0.9306	0.9051	0.8913	0.8837	0.8688	0.8543	0.8468	0.8307	0.8222	0.8158
$\alpha = 0.5, \lambda = 3$	0.9990	0.9675	0.9572	0.9385	0.9149	0.9059	0.9088	0.8915	0.8833	0.8642	0.8515	0.8437
$\alpha = 0.9, \lambda = 3$	1.0204	0.7576	0.7246	0.9388	0.7419	0.7077	0.9375	0.7258	0.6923	0.8723	0.6949	0.6613
$\alpha = 0.1, \lambda = 5$	0.9512	0.9068	0.8926	0.8869	0.8411	0.8280	0.8570	0.8155	0.8028	0.8227	0.7847	0.7726
$\alpha = 0.5, \lambda = 5$	1.0021	0.9556	0.9397	0.9286	0.8686	0.8528	0.9026	0.8527	0.8380	0.8573	0.8159	0.8026
$\alpha = 0.9, \lambda = 5$	1.0000	0.7143	0.6250	1.0000	0.7143	0.6250	0.8000	0.5714	0.5000	0.8000	0.5714	0.5714

Table 8.I-10 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.5 for INAR(1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$	Croston $\alpha = 0.5$	SBA $\alpha = 0.5$	SBJ $A = 0.5$
$\alpha = 0.1, \lambda = 0.5$	0.8561	0.9604	0.9661	0.8196	0.9151	0.9187	0.7849	0.8858	0.8931	0.7399	0.8464	0.8567
$\alpha = 0.5, \lambda = 0.5$	0.9043	1.0091	1.0141	0.8535	0.9527	0.9551	0.8062	0.9129	0.9190	0.7588	0.8601	0.8668
$\alpha = 0.9, \lambda = 0.5$	1.0597	0.8030	0.6621	1.0240	0.7841	0.6514	1.0063	0.7790	0.6469	0.9540	0.7399	0.6155
$\alpha = 0.1, \lambda = 1$	0.8545	0.9039	0.8707	0.7854	0.8350	0.8092	0.7494	0.8087	0.7858	0.7065	0.7673	0.7476
$\alpha = 0.5, \lambda = 1$	0.8989	0.9150	0.8729	0.8459	0.8734	0.8343	0.8072	0.8468	0.8139	0.7685	0.8075	0.7742
$\alpha = 0.9, \lambda = 1$	1.0649	0.6216	0.4873	1.0164	0.6048	0.4736	1.0082	0.6263	0.4909	0.9450	0.5712	0.4472
$\alpha = 0.1, \lambda = 3$	0.7493	0.6930	0.6140	0.7068	0.6516	0.5799	0.6679	0.6167	0.5491	0.6462	0.6053	0.5406
$\alpha = 0.5, \lambda = 3$	0.8756	0.7363	0.6295	0.8082	0.6997	0.6050	0.7869	0.6843	0.5917	0.7447	0.6508	0.5625
$\alpha = 0.9, \lambda = 3$	1.0417	0.3817	0.2874	1.0222	0.3770	0.2822	1.0000	0.3719	0.2795	0.9535	0.3596	0.2715
$\alpha = 0.1, \lambda = 5$	0.7364	0.6034	0.5083	0.6842	0.5666	0.4794	0.6585	0.5442	0.4584	0.6361	0.5287	0.4476
$\alpha = 0.5, \lambda = 5$	0.8716	0.6526	0.5337	0.7972	0.5955	0.4873	0.7758	0.5884	0.4812	0.7404	0.5633	0.4614
$\alpha = 0.9, \lambda = 5$	1.0000	0.2941	0.2273	1.0000	0.3125	0.2381	1.0000	0.2667	0.1905	1.0000	0.2667	0.2105

Table 8.I-11 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.2 for INARMA(1,1) series

Parameters	$n = 24$			$n = 36$			$n = 48$			$n = 96$		
	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$	Croston $\alpha = 0.2$	SBA $\alpha = 0.2$	SBJ $A = 0.2$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.9190	0.9654	0.9696	0.9551	0.9896	0.9920	0.9387	0.9724	0.9748	0.8993	0.9321	0.9346
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.0104	1.0541	1.0576	0.9574	0.9895	0.9918	0.9131	0.9504	0.9532	0.8746	0.9094	0.9120
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	1.0190	1.0438	1.0452	0.9688	0.9964	0.9980	0.9232	0.9541	0.9562	0.8675	0.9024	0.9051
$\alpha = 0.9, \beta = 0.1, \lambda = 0.5$	1.0060	0.9521	0.9398	0.9600	0.9113	0.9004	0.9176	0.8894	0.8806	0.8607	0.8460	0.8388
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.9892	1.0070	1.0065	0.8959	0.8533	0.8398	0.9082	0.9370	0.9377	0.8690	0.8946	0.8951
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0118	1.0259	1.0245	0.9382	0.9564	0.9557	0.9108	0.9328	0.9327	0.8610	0.8808	0.8805
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	1.0204	1.0394	1.0385	0.9611	0.9754	0.9741	0.9162	0.9332	0.9325	0.8625	0.8835	0.8833
$\alpha = 0.9, \beta = 0.1, \lambda = 1$	1.0111	0.9258	0.9037	0.9531	0.8750	0.8556	0.9201	0.8685	0.8520	0.8585	0.8062	0.7913
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.9550	0.9088	0.8948	0.8946	0.8574	0.8447	0.8537	0.8125	0.8000	0.8299	0.7949	0.7830
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9852	0.9053	0.8866	0.9208	0.8598	0.8429	0.8834	0.8362	0.8213	0.8365	0.7916	0.7778
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	1.0278	0.9487	0.9250	0.9444	0.8718	0.8718	0.9167	0.8462	0.8250	0.8571	0.8108	0.7895
$\alpha = 0.9, \beta = 0.1, \lambda = 5$	1.0000	0.6250	0.6250	1.0000	0.7143	0.6250	0.8000	0.5714	0.5000	0.8000	0.5714	0.5714

Table 8.I-12 $\mathrm{MASE_{INARMA}} / \mathrm{MASE_{Benchmark}}$ of lead-time forecasts $(l = 6)$ with smoothing parameter 0.5 for INARMA(1,1) series

Parameters	$n=24$			$n=36$			$n=48$			$n=96$		
	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$	Croston $\alpha=0.5$	SBA $\alpha=0.5$	SBJ $A=0.5$
$\alpha=0.1, \beta=0.1, \lambda=0.5$	0.8436	0.9460	0.9554	0.8118	0.9176	0.9250	0.7733	0.8817	0.8922	0.7361	0.8400	0.8521
$\alpha=0.1, \beta=0.9, \lambda=0.5$	0.9077	1.0043	1.0022	0.8329	0.9281	0.9292	0.7716	0.8820	0.8923	0.7343	0.8368	0.8437
$\alpha=0.5, \beta=0.5, \lambda=0.5$	0.9502	1.0183	1.0071	0.8870	0.9706	0.9628	0.8296	0.9157	0.9126	0.7871	0.8756	0.8742
$\alpha=0.9, \beta=0.1, \lambda=0.5$	1.0448	0.7635	0.6331	1.0426	0.7649	0.6361	0.9998	0.7685	0.6390	0.9469	0.7337	0.6099
$\alpha=0.1, \beta=0.1, \lambda=1$	0.8417	0.8799	0.8498	0.6944	0.5712	0.4790	0.7451	0.8140	0.7957	0.7078	0.7706	0.7513
$\alpha=0.1, \beta=0.9, \lambda=1$	0.8727	0.9003	0.8607	0.7934	0.8331	0.8001	0.7713	0.8107	0.7805	0.7290	0.7629	0.7318
$\alpha=0.5, \beta=0.5, \lambda=1$	0.9169	0.9273	0.8766	0.8656	0.8799	0.8314	0.8179	0.8400	0.7997	0.7738	0.7972	0.7581
$\alpha=0.9, \beta=0.1, \lambda=1$	1.0504	0.6051	0.4690	1.0216	0.6003	0.4673	0.9949	0.6140	0.4789	0.9411	0.5804	0.4538
$\alpha=0.1, \beta=0.1, \lambda=5$	0.7387	0.6121	0.5152	0.6994	0.5795	0.4883	0.6619	0.5486	0.4636	0.6480	0.5389	0.4545
$\alpha=0.1, \beta=0.9, \lambda=5$	0.8082	0.5982	0.4868	0.7500	0.5670	0.4637	0.7260	0.5548	0.4544	0.6815	0.5287	0.4350
$\alpha=0.5, \beta=0.5, \lambda=5$	0.9024	0.6379	0.5139	0.8293	0.5965	0.4789	0.8049	0.5789	0.4714	0.7500	0.5556	0.4478
$\alpha=0.9, \beta=0.1, \lambda=5$	1.0000	0.2941	0.2174	1.0000	0.3125	0.2273	1.0000	0.2500	0.1905	1.0000	0.2667	0.2000

Appendix 9.A INARMA(0,0), INAR(1), INMA(1) and INARMA(1,1) Series of 16,000 Series

In this appendix, the identified INARMA series of 16,000 data set are separated and forecasted with the known INARMA models.

Investigating the estimated parameter of the INARMA(0,0) process ($\hat{\lambda}$), we found that in general $\hat{\lambda}$ is close to 0.1 (the average is 0.1953 and 69.26 percent are between 0 and 0.15).

The estimated autoregressive parameter of the INAR(1) process, $\hat{\alpha}$, is close to 0.2 (the average is 0.2460 and 52.94 percent are between 0.1 and 0.3) and the estimated innovation parameter, $\hat{\lambda}$, is around 0.5 (the average is 0.3562 and 97.06 percent are between 0 and 1).

Looking at the estimated parameter of the INMA(1) process ($\hat{\beta}, \hat{\lambda}$) reveals that in general, $\hat{\beta}$ is close to zero (the average is 0.0898 and 46.29 percent are between 0 and 0.1) and $\hat{\lambda}$ is around 0.3 (the average is 0.3782 and 55.56 percent are between 0.2 and 0.4).

In general, the estimated autoregressive parameter of an INARMA(1,1) process is in the range $0.1 < \hat{\alpha} < 0.3$ (the average is 0.2988 and 66.67 percent are between 0.05 and 0.35), the moving average parameter, $\hat{\beta}$, is close to zero (the average is 0.1405 and 77.78 percent are between 0 and 0.1) and the innovation parameter, $\hat{\lambda}$, is around 0.3 (the average is 0.3558 and 44.44 percent are between 0.2 and 0.5).

Table 9.A-1 Only INARMA(0,0) series for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0712	-0.0922	-0.0399	-0.0088	-0.0364	0.0190	0.0340
MSE	0.3842	0.4127	0.3735	0.3789	0.3727	0.3777	0.3458
MASE	2.8850	2.9105	2.7293	2.5143	2.7120	2.3846	1.8711
PB of MASE (INARMA/Benchmark)	0.6661	0.6833	0.6099	0.5447	0.6030	0.4931	-
RGRMSE (INARMA/Benchmark)	0.8000	0.7595	0.8511	0.9314	0.8593	1.0254	-

Table 9.A-2 Only INARMA(0,0) series for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0271	-0.0443	0.0019	0.0325	0.0051	0.0580	0.0381
MSE	0.5258	0.5584	0.5165	0.5264	0.5160	0.5278	0.4935
MASE	0.3255	0.3321	0.3135	0.2997	0.3122	0.2900	0.2963
PB of MASE (INARMA/Benchmark)	0.5592	0.5266	0.4927	0.3845	0.4841	0.3453	-
RGRMSE (INARMA/Benchmark)	0.9224	0.9481	0.9998	1.1890	1.0092	1.3078	-

Table 9.A-3 Only INAR(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.1172	-0.1616	-0.0521	0.0122	-0.0449	0.0701	0.0386	0.0422	0.0457
MSE	0.9495	1.0586	0.9230	0.9502	0.9209	0.9398	0.8827	0.8542	0.8452
MASE	2.2362	2.2869	2.1174	1.9721	2.1046	1.8764	1.5807	1.5443	1.5641
PB of MASE INAR(1)-CLS/Benchmark	0.6266	0.5792	0.6176	0.5204	0.6176	0.4992			
PB of MASE INAR(1)-YW/Benchmark	0.6258	0.5784	0.6119	0.5155	0.6111	0.4935			
PB of MASE INAR(1)-CML/Benchmark	0.5866	0.5931	0.5613	0.5474	0.5605	0.5172			
RGRMSE INAR(1)-CLS/Benchmark	0.8077	0.8523	0.8349	0.9973	0.8367	1.0522			
RGRMSE INAR(1)-YW/Benchmark	0.8150	0.8625	0.8432	1.0122	0.8453	1.0686			
RGRMSE INAR(1)-CML/Benchmark	0.8293	0.8930	0.8574	1.0509	0.8606	1.1124			

Table 9.A-4 Only INAR(1) series for issue points (16000 series)

Accuracy measure	Forecasting method								
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INAR(1) CLS	INAR(1) YW	INAR(1) CML
ME	-0.0021	-0.0407	0.0522	0.1049	0.0583	0.1534	-0.2944	-0.2227	-0.1700
MSE	0.9913	1.1056	0.9928	1.0602	0.9938	1.0695	1.2621	1.1209	1.1887
MASE	0.7301	0.7376	0.7083	0.6813	0.7061	0.6661	1.1370	0.9934	1.0909
PB of MASE INAR(1)-CLS/Benchmark	0.2832	0.3322	0.2802	0.3027	0.2770	0.2900			
PB of MASE INAR(1)-YW/Benchmark	0.2733	0.3346	0.2719	0.2987	0.2729	0.2926			
PB of MASE INAR(1)-CML/Benchmark	0.3534	0.3714	0.3077	0.3391	0.3112	0.3355			
RGRMSE INAR(1)-CLS/Benchmark	2.0495	3.1778	2.2240	4.0306	2.2417	4.4597			
RGRMSE INAR(1)-YW/Benchmark	1.7792	2.6921	1.9248	3.3890	1.9391	3.7394			
RGRMSE INAR(1)-CML/Benchmark	1.9025	2.9755	2.0624	3.7819	2.0803	4.1872			

Table 9.A-5 Only INMA(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Crosto n 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	-0.0892	-0.1162	-0.0398	0.0139	-0.0343	0.0573	-0.0156	-0.0131
MSE	0.5758	0.6249	0.5654	0.5831	0.5648	0.5844	0.5606	0.5608
MASE	1.0865	1.0742	1.0350	0.9461	1.0294	0.9060	0.8965	0.8895
PB of MASE INMA(1)-CLS/Benchmark	0.6728	0.6260	0.6240	0.5072	0.6157	0.4676		
PB of MASE INMA(1)-YW/Benchmark	0.6728	0.6312	0.6173	0.5103	0.6096	0.4702		
RGRMSE INMA(1)-CLS/Benchmark	0.8180	0.8508	0.8718	0.9943	0.8783	1.0747		
RGRMSE INMA(1)-YW/Benchmark	0.8170	0.8481	0.8702	0.9903	0.8767	1.0698		

Table 9.A-6 Only INMA(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Crosto n 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INMA(1) CLS	INMA(1) YW
ME	0.0159	-0.0170	0.0651	0.1142	0.0705	0.1580	-0.0273	-0.0225
MSE	0.7990	0.8695	0.7966	0.8381	0.7969	0.8456	0.8234	0.8195
MASE	0.5068	0.5157	0.4936	0.4773	0.4923	0.4670	0.5560	0.5578
PB of MASE INMA(1)-CLS/Benchmark	0.4040	0.4371	0.3716	0.3423	0.3671	0.3181		
PB of MASE INMA(1)-YW/Benchmark	0.4155	0.4375	0.3714	0.3451	0.3699	0.3174		
RGRMSE INMA(1)-CLS/Benchmark	1.1810	1.3682	1.2527	1.6266	1.2615	1.7713		
RGRMSE INMA(1)-YW/Benchmark	1.1956	1.4012	1.2687	1.6694	1.2777	1.8191		

Table 9.A-7 Only INARMA(1,1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Crosto n 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.2371	-0.2913	-0.1649	-0.0973	-0.1569	-0.0327	-0.0978	-0.0634
MSE	1.0050	1.1561	0.9620	1.0136	0.9582	0.9916	1.1628	1.4696
MASE	1.4222	1.7511	1.3342	1.4445	1.3244	1.3477	0.9548	0.9417
PB of MASE INMA(1)-CLS/Benchmark	0.6574	0.6235	0.6173	0.5185	0.6080	0.5247		
PB of MASE INMA(1)-YW/Benchmark	0.6173	0.6235	0.6142	0.5833	0.6142	0.5741		
RGRMSE INMA(1)-CLS/Benchmark	0.7437	0.7688	0.7764	0.9020	0.7927	0.9490		
RGRMSE INMA(1)-YW/Benchmark	0.9560	0.9756	0.9769	1.0442	0.9795	1.0778		

Table 9.A-8 Only INARMA(1,1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA CLS	INARMA YW
ME	-0.1156	-0.1086	-0.0425	0.0723	-0.0344	0.1326	-0.5009	-0.7056
MSE	0.9332	1.0443	0.8990	0.9502	0.8962	0.9452	1.8226	2.4948
MASE	0.7472	0.7715	0.7244	0.7091	0.7219	0.6918	1.0096	1.1938
PB of MASE INMA(1)-CLS/Benchmark	0.3510	0.3420	0.3439	0.3179	0.3386	0.3346		
PB of MASE INMA(1)-YW/Benchmark	0.2706	0.3102	0.2653	0.2626	0.2653	0.2721		
RGRMSE INMA(1)-CLS/Benchmark	1.1273	1.2343	1.1858	1.4311	1.1958	1.5188		
RGRMSE INMA(1)-YW/Benchmark	1.1602	1.3154	1.3021	1.6644	1.2074	1.6381		

Appendix 9.B *h*-step-ahead Forecasts for the INARMA(0,0), INMA(1), and INARMA(1,1) Series

In this appendix, the results of three-step and six-step-ahead forecasts for INARMA(0,0), INMA(1), and INARMA(1,1) series of 16,000 and 3,000 series are presented. We only use YW-based forecasts for INMA(1) and CLS-based forecasts for INARMA(1,1) processes because of their better performance shown in chapter 9.

Table 9.B-1 Three-step-ahead INARMA(0,0) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0746	-0.0951	-0.0432	-0.0115	-0.0397	0.0163	0.0322
MSE	0.3858	0.4136	0.3745	0.3785	0.3736	0.3769	0.3452
MASE	2.8622	2.8788	2.7007	2.4709	2.6828	2.3373	1.8028
PB of MASE INMA(1)/Benchmark	0.6650	0.6817	0.6099	0.5442	0.6029	0.4922	-
RGRMSE INMA(1)/Benchmark	0.7974	0.7596	0.8490	0.9333	0.8575	1.0260	-

Table 9.B-2 Three-step-ahead INARMA(0,0) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0543	-0.0708	-0.0252	0.0061	-0.0220	0.0318	0.0115
MSE	0.4982	0.5347	0.4871	0.4969	0.4863	0.4967	0.4623
MASE	0.3070	0.3144	0.2940	0.2794	0.2926	0.2689	0.2758
PB of MASE INMA(1)/Benchmark	0.5636	0.5309	0.4944	0.3811	0.4846	0.3372	-
RGRMSE INMA(1)/Benchmark	0.9198	0.9481	0.9999	1.1991	1.0100	1.3212	-

Table 9.B-3 Six-step-ahead INARMA(0,0) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0776	-0.0974	-0.0463	-0.0141	-0.0428	0.0136	0.0292
MSE	0.3884	0.4189	0.3765	0.3813	0.3755	0.3789	0.3433
MASE	2.7616	2.7789	2.6008	2.3709	2.5830	2.2377	1.6869
PB of MASE INMA(1)/Benchmark	0.6585	0.6733	0.6041	0.5374	0.5975	0.4869	-
RGRMSE INMA(1)/Benchmark	3.1734	0.7644	0.8500	0.9379	0.8582	1.0284	-

Table 9.B-4 Six-step-ahead INARMA(0,0) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0565	-0.0725	-0.0270	0.0054	-0.0237	0.0314	0.0099
MSE	0.5233	0.5656	0.5120	0.5251	0.5112	0.5246	0.4876
MASE	0.3126	0.3193	0.2994	0.2838	0.2979	0.2731	0.2802
PB of MASE INMA(1)/Benchmark	0.5582	0.5266	0.4890	0.3766	0.4799	0.3353	-
RGRMSE INMA(1)/Benchmark	48.0891	0.9588	1.0071	1.2114	1.0168	1.3330	-

Table 9.B-5 Three-step-ahead YW-INMA(1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0961	-0.1218	-0.0465	0.0086	-0.0410	0.0521	-0.0199
MSE	0.5648	0.6122	0.5520	0.5652	0.5511	0.5649	0.5430
MASE	1.0845	1.0698	1.0333	0.9429	1.0277	0.9035	0.9096
PB of MASE INMA(1)/Benchmark	0.6716	0.6111	0.5931	0.4940	0.5801	0.4439	-
RGRMSE INMA(1)/Benchmark	0.8495	0.8765	0.9000	1.0432	0.9049	1.1251	-

Table 9.B-6 Three-step-ahead YW-INMA(1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0537	-0.0845	-0.0045	0.0461	0.0009	0.0897	-0.0254
MSE	0.6436	0.6859	0.6379	0.6542	0.6379	0.6615	0.6450
MASE	0.4549	0.4584	0.4405	0.4188	0.4390	0.4090	0.4510
PB of MASE INMA(1)/Benchmark	0.5124	0.4501	0.4134	0.3402	0.3993	0.3014	-
RGRMSE INMA(1)/Benchmark	1.0113	1.1146	1.0770	1.3532	1.0835	1.4670	-

Table 9.B-7 Six-step-ahead YW-INMA(1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1062	-0.1325	-0.0562	-0.0008	-0.0506	0.0431	-0.0258
MSE	0.5831	0.6348	0.5669	0.5774	0.5657	0.5743	0.5416
MASE	1.0903	1.0788	1.0365	0.9449	1.0306	0.9030	0.8978
PB of MASE INMA(1)/Benchmark	0.6750	0.6332	0.6147	0.5143	0.6004	0.4648	-
RGRMSE INMA(1)/Benchmark	0.8323	0.8520	0.8856	1.0289	0.8904	1.1094	-

Table 9.B-8 Six-step-ahead YW-INMA(1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0778	-0.1058	-0.0284	0.0248	-0.0229	0.0683	-0.0469
MSE	0.6393	0.6936	0.6251	0.6377	0.6242	0.6378	0.6079
MASE	0.4570	0.4620	0.4403	0.4163	0.4385	0.4042	0.4411
PB of MASE INMA(1)/Benchmark	0.5472	0.4727	0.4492	0.3531	0.4317	0.3152	-
RGRMSE INMA(1)/Benchmark	0.9911	1.1070	1.0561	1.3687	1.0607	1.4705	-

Table 9.B-9 Three-step-ahead INARMA(1,1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2307	-0.2808	-0.1586	-0.0881	-0.1506	-0.0238	-0.1343
MSE	1.0281	1.0745	0.9868	0.9658	0.9831	0.9545	0.9229
MASE	1.4278	1.7203	1.3430	1.4245	1.3337	1.3324	0.9570
PB of MASE INMA(1)/Benchmark	0.5850	0.5752	0.5784	0.5033	0.5784	0.4837	-
RGRMSE INMA(1)/Benchmark	0.8806	0.8863	0.9032	1.0710	0.9070	1.1219	-

Table 9.B-10 Three-step-ahead INARMA(1,1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1683	-0.1578	-0.0956	0.0213	-0.0875	0.0810	-0.2303
MSE	1.1600	1.2204	1.1204	1.1269	1.1171	1.1220	1.1495
MASE	0.7783	0.8064	0.7483	0.7143	0.7452	0.6898	0.7602
PB of MASE INMA(1)/Benchmark	0.4300	0.4781	0.4406	0.3586	0.4406	0.3414	-
RGRMSE INMA(1)/Benchmark	1.0953	1.1663	1.1447	1.4514	1.1485	1.5328	-

Table 9.B-11 Six-step-ahead INARMA(1,1) for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2288	-0.2772	-0.1565	-0.0842	-0.1484	-0.0199	-0.1345
MSE	1.0917	1.1648	1.0484	1.0479	1.0445	1.0336	0.9695
MASE	1.4417	1.7401	1.3558	1.4414	1.3463	1.3488	0.9725
PB of MASE INMA(1)/Benchmark	0.5735	0.5663	0.5591	0.5054	0.5591	0.4803	-
RGRMSE INMA(1)/Benchmark	0.8870	0.9407	0.9210	1.0856	0.9549	1.1495	-

Table 9.B-12 Six-step-ahead INARMA(1,1) for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1851	-0.1649	-0.1126	0.0115	-0.1045	0.0702	-0.2529
MSE	1.0022	0.9567	0.9606	0.8924	0.9571	0.8962	1.0479
MASE	0.7526	0.7172	0.7256	0.6545	0.7226	0.6361	0.7774
PB of MASE INMA(1)/Benchmark	0.3955	0.4300	0.3887	0.3126	0.3887	0.2645	-
RGRMSE INMA(1)/Benchmark	1.0631	1.1909	1.1345	1.4448	1.1447	1.5675	-

Table 9.B-13 Three-step-ahead INARMA(0,0) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0073	-0.1147	0.1498	0.2722	0.1656	0.4011	0.1209
MSE	2.0963	2.3856	2.1016	2.2770	2.1054	2.3346	2.0757
MASE	0.9130	0.9829	0.9001	0.9306	0.8994	0.9315	0.8951
PB of MASE INMA(1)/Benchmark	0.5375	0.5658	0.5063	0.5387	0.5063	0.5441	-
RGRMSE INMA(1)/Benchmark	0.8833	0.8402	0.9123	0.9056	0.9147	0.9050	-

Table 9.B-14 Three-step-ahead INARMA(0,0) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0021	-0.1119	0.1385	0.2670	0.1541	0.3933	0.0733
MSE	2.2061	2.5051	2.2064	2.3882	2.2096	2.4435	2.1745
MASE	0.9435	1.0138	0.9290	0.9580	0.9281	0.9576	0.9252
PB of MASE INMA(1)/Benchmark	0.5361	0.5703	0.5024	0.5350	0.5024	0.5369	-
RGRMSE INMA(1)/Benchmark	0.9236	0.8874	0.9698	0.9618	0.9752	0.9654	-

Table 9.B-15 Six-step-ahead INARMA(0,0) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0082	-0.1489	0.1333	0.2400	0.1490	0.3696	0.1054
MSE	2.2315	2.6177	2.2292	2.4492	2.2323	2.4920	2.1945
MASE	0.9165	0.9959	0.9025	0.9384	0.9016	0.9376	0.8979
PB of MASE INMA(1)/Benchmark	0.5435	0.5729	0.5046	0.5508	0.5044	0.5493	-
RGRMSE INMA(1)/Benchmark	0.9328	0.8844	0.9559	0.9425	0.9545	0.9412	-

Table 9.B-16 Six-step-ahead INARMA(0,0) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0172	-0.1033	0.1564	0.2749	0.1719	0.4010	0.0902
MSE	2.3879	2.7882	2.3937	2.6505	2.3978	2.7048	2.3370
MASE	0.9325	1.0139	0.9191	0.9599	0.9183	0.9603	0.9151
PB of MASE INMA(1)/Benchmark	0.5422	0.5749	0.4998	0.5521	0.5024	0.5505	-
RGRMSE INMA(1)/Benchmark	0.9801	0.9327	1.0119	0.9919	1.0037	0.9906	-

Table 9.B-17 Three-step-ahead YW-INMA(1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1216	-0.1155	0.1574	0.5804	0.1884	0.8124	-0.1302
MSE	4.4836	5.2211	4.4409	5.1449	4.4485	5.4210	4.3595
MASE	0.6737	0.7185	0.6584	0.6887	0.6576	0.6979	0.6644
PB of MASE INMA(1)/Benchmark	0.5140	0.5544	0.4862	0.5215	0.4888	0.5269	-
RGRMSE INMA(1)/Benchmark	0.9363	0.9402	0.9845	1.0530	0.9921	1.0364	-

Table 9.B-18 Three-step-ahead YW-INMA(1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1423	-0.1298	0.1346	0.5594	0.1654	0.7891	-0.1836
MSE	4.5829	5.3247	4.5234	5.2145	4.5292	5.4785	4.4694
MASE	0.7248	0.7702	0.7079	0.7375	0.7069	0.7469	0.7187
PB of MASE INMA(1)/Benchmark	0.5041	0.5456	0.4744	0.5123	0.4760	0.5196	-
RGRMSE INMA(1)/Benchmark	0.9563	0.9815	1.0174	1.1065	1.0247	1.0818	-

Table 9.B-19 Six-step-ahead YW-INMA(1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2020	-0.2161	0.0805	0.4937	0.1119	0.7303	-0.1898
MSE	4.6505	5.5099	4.5500	5.2474	4.5521	5.4868	4.5450
MASE	0.6775	0.7286	0.6572	0.6837	0.6558	0.6881	0.6668
PB of MASE INMA(1)/Benchmark	0.5215	0.5620	0.4732	0.5153	0.4732	0.5129	-
RGRMSE INMA(1)/Benchmark	0.9514	0.9522	1.0185	1.1714	1.0252	1.1514	-

Table 9.B-20 Six-step-ahead YW-INMA(1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2165	-0.2140	0.0634	0.4851	0.0945	0.7181	-0.2423
MSE	4.7246	5.5844	4.6131	5.3183	4.6140	5.5557	4.6174
MASE	0.7182	0.7704	0.6966	0.7229	0.6951	0.7271	0.7093
PB of MASE INMA(1)/Benchmark	0.5164	0.5565	0.4665	0.5055	0.4647	0.5045	-
RGRMSE INMA(1)/Benchmark	0.9726	1.0051	1.0493	1.2585	1.0565	1.2503	-

Table 9.B-21 Three-step-ahead CLS-INARMA(1,1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1581	-0.1234	0.1331	0.5960	0.1655	0.8358	-0.2299
MSE	5.0346	5.7062	4.9548	5.6299	4.9593	5.9228	5.0253
MASE	0.8851	0.9393	0.8715	0.9113	0.8710	0.9247	0.8922
PB of MASE INARMA/Benchmark	0.4943	0.5371	0.4686	0.4971	0.4557	0.5286	-
RGRMSE INARMA/Benchmark	1.0870	1.0434	1.0891	1.0884	1.1024	1.1096	-

Table 9.B-22 Three-step-ahead CLS-INARMA(1,1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1527	-0.1184	0.1367	0.5966	0.1689	0.8349	-0.2622
MSE	5.1838	5.8418	5.0951	5.7476	5.0987	6.0369	5.2148
MASE	0.9129	0.9683	0.9009	0.9436	0.9006	0.9587	0.9263
PB of MASE INARMA/Benchmark	0.4802	0.5268	0.4567	0.5027	0.4419	0.5292	-
RGRMSE INARMA/Benchmark	1.1165	1.0876	1.1226	1.1171	1.1361	1.1532	-

Table 9.B-23 Six-step-ahead CLS-INARMA(1,1) for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2078	-0.1735	0.0885	0.5586	0.1214	0.8027	-0.2570
MSE	5.1627	6.1418	5.0427	5.9093	5.0439	6.1776	5.2268
MASE	0.8862	0.9431	0.8615	0.8986	0.8600	0.9045	0.8915
PB of MASE INARMA/Benchmark	0.4878	0.5163	0.4531	0.5020	0.4592	0.4980	-
RGRMSE INARMA/Benchmark	1.0359	1.0342	1.1680	1.1983	1.1649	1.2362	-

Table 9.B-24 Six-step-ahead CLS-INARMA(1,1) for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2145	-0.1673	0.0793	0.5555	0.1120	0.7964	-0.3096
MSE	5.4890	6.4631	5.3563	6.2278	5.3563	6.4971	5.6144
MASE	0.9180	0.9736	0.8919	0.9263	0.8903	0.9319	0.9285
PB of MASE INARMA/Benchmark	0.4748	0.5088	0.4422	0.4890	0.4459	0.4863	-
RGRMSE INARMA/Benchmark	1.0322	1.0729	1.1896	1.2521	1.1945	1.3254	-

Appendix 9.C Lead Time Forecasts for the INARMA(0,0), INMA(1), and INARMA(1,1) Series

In this appendix, the results of lead time forecasting for INARMA(0,0), INMA(1) and INARMA(1,1) processes of 16,000 and 3,000 series are presented. Two values have been assumed for the lead time: $l = 3, 6$. The INARMA lead-time forecasts based on the Equation 9-7 results have also been compared to the cumulative h-step ahead forecasts over lead time.

Table 9.C-1 Lead-time forecasts $(l = 3)$ for INARMA(0,0) series for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.2255	-0.2873	-0.1314	-0.0364	-0.1209	0.0472	0.1003
MSE	1.5975	1.8564	1.4946	1.5350	1.4866	1.5190	1.2385
MASE	7.0789	7.0809	6.6157	5.9388	6.5653	5.5852	4.5738
PB of MASE (INARMA/Benchmark)	0.6131	0.6328	0.5799	0.5410	0.5758	0.5097	-
RGRMSE (INARMA/Benchmark)	10.6388	0.7858	0.8691	0.9211	0.8729	0.9808	-

Table 9.C-2 Lead-time forecasts $(l = 3)$ for INARMA(0,0) series for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.1426	-0.1921	-0.0552	0.0388	-0.0455	0.1158	0.1202
MSE	2.0008	2.3265	1.9029	1.9933	1.8961	1.9940	1.6934
MASE	0.6945	0.7207	0.6681	0.6457	0.6655	0.6282	0.6048
PB of MASE (INARMA/Benchmark)	0.6328	0.6198	0.5956	0.5120	0.5914	0.4746	-
RGRMSE (INARMA/Benchmark)	56.9213	0.8351	0.8725	0.9870	0.8742	1.0459	-

Table 9.C-3 Lead-time forecasts $(l = 6)$ for INARMA(0,0) series for all points in time (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.4710	-0.5895	-0.2830	-0.0901	-0.2621	0.0764	0.1955
MSE	4.5190	5.5680	4.1059	4.2688	4.0734	4.1985	3.1290
MASE	12.1123	12.0836	11.2398	9.9069	11.1454	9.2546	7.3915
PB of MASE (INARMA/Benchmark)	0.5802	0.5932	0.5559	0.5338	0.5525	0.5141	-
RGRMSE (INARMA/Benchmark)	1.1190	0.7957	0.8778	0.9139	1.1686	0.9898	-

Table 9.C-4 Lead-time forecasts $(l = 6)$ for INARMA(0,0) series for issue points (16000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.3379	-0.4337	-0.1605	0.0336	-0.1408	0.1894	0.2047
MSE	5.3925	6.8288	5.0226	5.4871	4.9985	5.5072	4.4452
MASE	1.1360	1.2038	1.0880	1.0623	1.0834	1.0357	0.9957
PB of MASE (INARMA/Benchmark)	0.5746	0.5631	0.5476	0.4892	0.5443	0.4718	-
RGRMSE (INARMA/Benchmark)	0.9209	0.9066	0.9738	1.0623	0.9777	1.1819	-

Table 9.C-5 Lead-time forecasts $(l = 3)$ for INMA(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.2834	-0.3605	-0.1346	0.0307	-0.1181	0.1611	-0.0544	-0.0554
MSE	2.1828	2.6271	2.0754	2.2196	2.0685	2.2225	1.9954	2.0142
MASE	2.3743	2.3692	2.2585	2.0808	2.2466	2.0062	1.9328	1.9081
PB of MASE INARMA-LT/Benchmark	0.6334	0.6035	0.5773	0.5289	0.5730	0.4995		
PB of MASE INARMA-h/Benchmark	0.6334	0.6002	0.5719	0.5272	0.5664	0.4946		
RGRMSE INARMA-LT/Benchmark	0.8106	0.8214	0.8647	0.9050	0.8771	0.9681		
RGRMSE INARMA-h/Benchmark	0.8802	0.8919	0.9422	0.9772	0.9530	1.0604		

Table 9.C-6 Lead-time forecasts $(l = 3)$ for INMA(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.1116	-0.2041	0.0359	0.1880	0.0523	0.3186	-0.0298	0.0098
MSE	2.4767	3.0064	2.4170	2.6614	2.4160	2.7077	2.4746	2.4833
MASE	0.9346	1.0050	0.9084	0.9197	0.9062	0.9088	0.9227	0.9479
PB of MASE INARMA-LT/Benchmark	0.5177	0.5462	0.4638	0.4661	0.4643	0.4503		
PB of MASE INARMA-h/Benchmark	0.5754	0.5417	0.4932	0.4894	0.4865	0.4667		
RGRMSE INARMA-LT/Benchmark	0.9244	0.9666	0.9867	1.0462	1.0107	1.1183		
RGRMSE INARMA-h/Benchmark	0.9495	0.9963	1.0247	1.0605	1.0454	1.1379		

Table 9.C-7 Lead-time forecasts $(l=6)$ for INMA(1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.5984	-0.7560	-0.2981	0.0341	-0.2648	0.2975	-0.1156	-0.1205
MSE	6.1055	7.9654	5.5954	6.0886	5.5593	6.0385	4.9485	5.0550
MASE	3.9268	3.9254	3.6929	3.3373	3.6697	3.2155	3.0185	3.2643
PB of MASE INARMA-LT/Benchmark	0.6272	0.6045	0.5854	0.5376	0.5789	0.5203		
PB of MASE INARMA-h/Benchmark	0.6207	0.6039	0.5651	0.5221	0.5621	0.5149		
RGRMSE INARMA-LT/Benchmark	15.2149	0.7831	0.8800	0.9654	0.8817	0.9951		
RGRMSE INARMA-h/Benchmark	15.3080	0.8678	0.9714	1.0523	0.9740	1.0662		

Table 9.C-8 Lead-time forecasts $(l=6)$ for INMA(1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.3591	-0.5271	-0.0624	0.2564	-0.0295	0.5176	-0.1806	-0.0805
MSE	6.6701	8.8973	6.2622	7.0723	6.2406	7.1395	6.0526	6.1128
MASE	1.4716	1.6277	1.4074	1.4446	1.4024	1.4285	1.4132	1.4973
PB of MASE INARMA-LT/Benchmark	0.5173	0.5313	0.4771	0.4678	0.4688	0.4824		
PB of MASE INARMA-h/Benchmark	0.5517	0.5336	0.4977	0.4735	0.5024	0.5077		
RGRMSE INARMA-LT/Benchmark	1.0678	0.9865	1.0592	1.0985	1.0584	1.1128		
RGRMSE INARMA-h/Benchmark	1.0667	0.9891	1.0741	1.0906	1.0721	1.1152		

Table 9.C-9 Lead-time forecasts $(l=3)$ for INARMA(1,1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.6985	-0.8490	-0.4823	-0.2708	-0.4583	-0.0780	-0.3636	-0.4199
MSE	3.9379	4.8160	3.5558	3.6977	3.5220	3.5494	4.5303	2.9121
MASE	3.4416	4.4321	3.1996	3.5549	3.1736	3.2997	2.1659	1.8775
PB of MASE INARMA-LT/Benchmark	0.5980	0.6144	0.5490	0.5098	0.5490	0.5163		
PB of MASE INARMA-h/Benchmark	0.6275	0.6111	0.6078	0.5425	0.6013	0.5261		
RGRMSE INARMA-LT/Benchmark	0.9087	0.8200	0.9427	1.0951	0.9697	1.1234		
RGRMSE INARMA-h/Benchmark	0.8273	0.7576	0.8615	1.0057	0.8866	1.0275		

Table 9.C-10 Lead-time forecasts $(l = 3)$ for INARMA(1,1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.4052	-0.3736	-0.1871	0.1637	-0.1629	0.3428	-1.2448	-0.5621
MSE	3.7800	4.6006	3.4474	3.7471	3.4202	3.6994	9.9470	3.7745
MASE	1.5416	1.7256	1.4707	1.5290	1.4640	1.4811	2.1597	1.6225
PB of MASE INARMA-LT/Benchmark	0.3630	0.3927	0.3459	0.3388	0.3516	0.3430		
PB of MASE INARMA-h/Benchmark	0.4902	0.5031	0.4820	0.4404	0.4820	0.4262		
RGRMSE INARMA-LT/Benchmark	1.7307	1.5257	1.8607	1.7542	1.9733	2.1105		
RGRMSE INARMA-h/Benchmark	1.3178	1.1992	1.4541	1.3574	1.5426	1.6203		

Table 9.C-11 Lead-time forecasts $(l = 6)$ for INARMA(1,1) series for all points in time (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.4841	-1.7743	-1.0500	-0.6165	-1.0018	-0.2306	-0.8883	-0.9773
MSE	11.1344	13.9431	9.4682	9.4214	9.3179	8.8019	11.7047	6.9610
MASE	5.9961	8.0869	5.4768	6.2301	5.4215	5.6872	3.5749	2.2767
PB of MASE INARMA-LT/Benchmark	0.5520	0.5627	0.5341	0.5376	0.5269	0.5269		
PB of MASE INARMA-h/Benchmark	0.5412	0.5771	0.5412	0.5233	0.5448	0.5197		
RGRMSE INARMA-LT/Benchmark	0.9543	0.9909	1.2085	1.2125	1.2981	1.2143		
RGRMSE INARMA-h/Benchmark	1.0306	1.1248	1.3587	1.3553	1.4787	1.3106		

Table 9.C-12 Lead-time forecasts $(l = 6)$ for INARMA(1,1) series for issue points (16000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.9955	-0.8740	-0.5601	0.1841	-0.5117	0.5368	-2.3703	-1.3726
MSE	10.7433	11.7804	9.2238	8.7355	9.0943	8.6310	31.9221	10.2158
MASE	2.4007	2.6007	2.2126	2.2321	2.1967	2.1917	3.4978	1.9589
PB of MASE INARMA-LT/Benchmark	0.3889	0.4205	0.3122	0.3492	0.3122	0.3223		
PB of MASE INARMA-h/Benchmark	0.3838	0.4823	0.4180	0.4117	0.4359	0.3987		
RGRMSE INARMA-LT/Benchmark	1.4795	1.8284	1.8921	2.0133	1.9238	1.9300		
RGRMSE INARMA-h/Benchmark	1.2300	1.5394	1.5689	1.7109	1.5787	1.6232		

Table 9.C-13 Lead-time forecasts $(l = 3)$ for INARMA(0,0) series for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0580	-0.3078	0.4856	0.8527	0.5331	1.2396	0.4706
MSE	7.5489	10.1828	7.6207	9.2597	7.6575	9.7960	7.7135
MASE	1.7359	2.0008	1.7192	1.8872	1.7204	1.9243	1.7258
PB of MASE (INARMA/Benchmark)	0.4997	0.5589	0.5024	0.5665	0.5005	0.5849	-
RGRMSE (INARMA/Benchmark)	0.9288	0.8503	0.9570	0.9071	0.9593	0.8969	-

Table 9.C-14 Lead-time forecasts $(l = 3)$ for INARMA(0,0) series for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0239	-0.3534	0.3979	0.7833	0.4447	1.1622	0.3695
MSE	7.8662	10.6695	7.8380	9.5142	7.8642	9.9767	7.9653
MASE	1.7686	2.0423	1.7422	1.9105	1.7423	1.9416	1.7522
PB of MASE (INARMA/Benchmark)	0.4992	0.5580	0.4998	0.5586	0.4967	0.5789	-
RGRMSE (INARMA/Benchmark)	0.9951	0.9336	1.0350	1.0059	1.0420	1.0026	-

Table 9.C-15 Lead-time forecasts $(l = 6)$ for INARMA(0,0) series for all points in time (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	0.0938	-0.7503	0.9428	1.5832	1.0371	2.3610	1.2740
MSE	20.7260	32.6198	20.8990	27.6901	21.0370	29.6099	23.6958
MASE	2.7983	3.4401	2.7867	3.2239	2.7924	3.3257	2.9096
PB of MASE (INARMA/Benchmark)	0.4888	0.5616	0.4799	0.5588	0.4823	0.5768	-
RGRMSE (INARMA/Benchmark)	1.1199	1.0411	1.1375	1.0574	1.1083	1.0703	-

Table 9.C-16 Lead-time forecasts $(l = 6)$ for INARMA(0,0) series for issue points (3000 series)

Accuracy measure	Forecasting method						
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA
ME	-0.0025	-0.7255	0.8330	1.5439	0.9258	2.3004	1.0574
MSE	22.3042	34.7526	22.1901	29.4718	22.3007	31.3196	24.7486
MASE	2.8816	3.5402	2.8455	3.2920	2.8483	3.3783	2.9296
PB of MASE (INARMA/Benchmark)	0.4986	0.5698	0.4820	0.5594	0.4849	0.5783	-
RGRMSE (INARMA/Benchmark)	1.1981	1.0904	1.2219	1.1314	1.2145	1.1387	-

Table 9.C-17 Lead-time forecasts $(l = 3)$ for INMA(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.3688	-0.3504	0.4681	1.7373	0.5611	2.4332	-0.3874	-0.4181
MSE	17.5351	23.8375	17.1759	23.2984	17.2471	25.8320	16.7291	16.9991
MASE	1.3331	1.5167	1.2915	1.4657	1.2908	1.5342	1.3014	1.4582
PB of MASE INARMA-LT/Benchmark	0.5275	0.5662	0.4788	0.5593	0.4819	0.5754		
PB of MASE INARMA-h/Benchmark	0.5123	0.5570	0.4691	0.5481	0.4699	0.5728		
RGRMSE INARMA-LT/Benchmark	0.9531	0.9737	1.0577	1.0357	1.0625	1.0071		
RGRMSE INARMA-h/Benchmark	1.0479	1.0633	1.1618	1.1314	1.1611	1.0966		

Table 9.C-18 Lead-time forecasts $(l = 3)$ for INMA(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.3976	-0.3602	0.4332	1.7075	0.5255	2.3968	-0.5183	-0.5030
MSE	17.9749	24.1636	17.5602	23.5958	17.6255	26.1131	17.3941	17.6328
MASE	1.4317	1.6228	1.3859	1.5654	1.3852	1.6363	1.4137	1.5501
PB of MASE INARMA-LT/Benchmark	0.5102	0.5571	0.4669	0.5488	0.4690	0.5639		
PB of MASE INARMA-h/Benchmark	0.5029	0.5501	0.4614	0.5424	0.4618	0.5630		
RGRMSE INARMA-LT/Benchmark	0.9925	1.0288	1.1491	1.1038	1.1275	1.0930		
RGRMSE INARMA-h/Benchmark	1.0723	1.1080	1.2388	1.1815	1.2071	1.1704		

Table 9.C-19 Lead-time forecasts $(l = 6)$ for INMA(1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.1294	-1.2137	0.5656	3.0449	0.7540	4.4645	-1.0404	-1.1130
MSE	50.1739	78.5673	46.7922	70.3372	46.8927	79.3655	46.2845	48.3477
MASE	2.1986	2.6846	2.0800	2.5260	2.0759	2.6662	2.1066	2.4075
PB of MASE INARMA-LT/Benchmark	0.5485	0.6046	0.4781	0.5714	0.4810	0.5898		
PB of MASE INARMA-h/Benchmark	0.5178	0.5849	0.4748	0.5682	0.4707	0.5763		
RGRMSE INARMA-LT/Benchmark	0.9938	0.9710	1.1675	1.1387	1.1794	1.2102		
RGRMSE INARMA-h/Benchmark	1.1167	1.0784	1.2921	1.2627	1.3078	1.3256		

Table 9.C-20 Lead-time forecasts $(l = 6)$ for INMA(1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.1595	-1.1449	0.5198	3.0496	0.7064	4.4478	-1.3074	-1.2839
MSE	51.1774	79.8211	47.4778	71.6060	47.5466	80.6072	47.9506	49.9924
MASE	2.3426	2.8731	2.2081	2.6919	2.2034	2.8375	2.2635	2.5534
PB of MASE INARMA-LT/Benchmark	0.5362	0.6014	0.4618	0.5659	0.4646	0.5862		
PB of MASE INARMA-h/Benchmark	0.5147	0.5824	0.4659	0.5681	0.4616	0.5745		
RGRMSE INARMA-LT/Benchmark	1.0391	1.0196	1.2601	1.2098	1.2488	1.2893		
RGRMSE INARMA-h/Benchmark	1.1678	1.1203	1.3896	1.3244	1.3742	1.3897		

Table 9.C-21 Lead-time forecasts $(l = 3)$ for INARMA(1,1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.5187	-0.4146	0.3551	1.7437	0.4521	2.4631	-0.8267	-0.7909
MSE	18.8627	24.5580	18.0579	23.7418	18.0889	26.3350	22.0895	19.7577
MASE	1.6389	1.8766	1.5921	1.8215	1.5923	1.9362	1.7349	1.6339
PB of MASE INARMA-LT/Benchmark	0.4614	0.5671	0.4271	0.5100	0.4457	0.5486		
PB of MASE INARMA-h/Benchmark	0.4914	0.5771	0.4414	0.5243	0.4357	0.5557		
RGRMSE INARMA-LT/Benchmark	1.0506	1.0450	1.1573	1.1289	1.1518	1.1153		
RGRMSE INARMA-h/Benchmark	1.0372	1.0547	1.1599	1.1571	1.1531	1.1100		

Table 9.C-22 Lead-time forecasts $(l = 3)$ for INARMA(1,1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-0.5784	-0.4755	0.2898	1.6694	0.3863	2.3843	-1.1568	-0.9061
MSE	19.6577	25.6292	18.6439	24.2308	18.6526	26.6505	23.2961	20.6115
MASE	1.7240	1.9795	1.6729	1.8928	1.6727	1.9998	1.8456	1.8000
PB of MASE INARMA-LT/Benchmark	0.4417	0.5638	0.4098	0.4925	0.4309	0.5297		
PB of MASE INARMA-h/Benchmark	0.4800	0.5718	0.4344	0.5145	0.4328	0.5452		
RGRMSE INARMA-LT/Benchmark	1.1151	1.1157	1.2715	1.2833	1.2528	1.2352		
RGRMSE INARMA-h/Benchmark	1.0510	1.0761	1.2019	1.2151	1.1831	1.1476		

Table 9.C-23 Lead-time forecasts $(l = 6)$ for INARMA(1,1) series for all points in time (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.5365	-1.3308	0.2412	3.0621	0.4388	4.5264	-2.2603	-2.0711
MSE	57.1789	83.5607	51.1221	73.0681	50.9719	82.0184	79.4808	65.2786
MASE	2.8246	3.4084	2.6916	3.3006	2.6883	3.5013	3.0694	2.7508
PB of MASE INARMA-LT/Benchmark	0.4510	0.5612	0.4163	0.5347	0.4286	0.5571		
PB of MASE INARMA-h/Benchmark	0.4878	0.5571	0.4469	0.5204	0.4490	0.5490		
RGRMSE INARMA-LT/Benchmark	1.2453	1.1376	1.3844	1.2097	1.3820	1.1728		
RGRMSE INARMA-h/Benchmark	1.2742	1.2320	1.4736	1.2600	1.4718	1.1923		

Table 9.C-24 Lead-time forecasts $(l = 6)$ for INARMA(1,1) series for issue points (3000 series)

Accuracy measure	Forecasting method							
	Croston 0.2	Croston 0.5	SBA 0.2	SBA 0.5	SBJ 0.2	SBJ 0.5	INARMA-LT	INARMA-h
ME	-1.6333	-1.3500	0.1298	2.9869	0.3257	4.4325	-2.8186	-2.3110
MSE	61.3450	88.6872	54.3713	76.6571	54.1273	85.1684	86.2956	69.9140
MASE	2.9996	3.6386	2.8464	3.4328	2.8394	3.6155	3.2911	2.9643
PB of MASE INARMA-LT/Benchmark	0.4504	0.5676	0.4086	0.5062	0.4185	0.5239		
PB of MASE INARMA-h/Benchmark	0.4853	0.5648	0.4502	0.5103	0.4541	0.5326		
RGRMSE INARMA-LT/Benchmark	1.2606	1.1517	1.4470	1.4508	1.3907	1.3271		
RGRMSE INARMA-h/Benchmark	1.2663	1.2116	1.4723	1.4535	1.4050	1.2944		