Forecast Horizon Aggregation in Integer Autoregressive Moving Average (INARMA) Models

Maryam Mohammadipour[1], John E. Boylan

Buckinghamshire New University, High Wycombe Campus, Queen Alexandra Road, High Wycombe, Buckinghamshire, HP11 2JZ, UK

Abstract

This paper addresses aggregation in integer autoregressive moving average (INARMA) models. Although aggregation in continuous-valued time series has been widely discussed, the same is not true for integer-valued time series. Forecast horizon aggregation is addressed in this paper. It is shown that the overlapping forecast horizon aggregation of an INARMA process results in an INARMA process. The conditional expected value of the aggregated process is also derived, for use in forecasting. A simulation experiment is conducted to assess the accuracy of the forecasts produced by the aggregation method and to compare it to the accuracy of cumulative *h*-step ahead forecasts over the forecasting horizon. The results of an empirical analysis are also provided.

Key words: Discrete time series, INARMA model, temporal aggregation, cross-sectional aggregation, forecast horizon aggregation, Yule-Walker estimation

1. INTRODUCTION

Time series aggregation is a widely discussed subject for continuous-valued time series. It goes back over 50 years [1] and since then many papers have considered different aspects of

[1] Corresponding author.
E-mail addresses: mohammadipour@bucks.ac.uk (M. Mohammadipour), john.boylan@bucks.ac.uk (J.E. Boylan).

aggregation for continuous-valued time series (see for example: [2-9]). Three types of aggregation have been identified in the literature which can be classified as: cross-sectional aggregation, temporal aggregation, and forecast horizon aggregation. The first class of aggregation produces forecasts based on aggregated data series and the other two produce forecasts based on aggregated periods.

Cross-sectional or contemporaneous aggregation is conducted across individual series 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.

Temporal aggregation, also called flow scheme, refers to aggregation in which a low frequency time series (e.g. annual) is derived from a high frequency time series (e.g. quarterly or monthly). 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 ($k = 12$).

Finally, forecast horizon aggregation refers to the case in which there is requirement for a forecast of the total value over a number of time periods 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: [10]).

With respect to temporal/forecast horizon aggregation, we must distinguish between overlapping and non-overlapping cases. In non-overlapping aggregation, the demand series are divided into consecutive non-overlapping blocks of equal length. In overlapping aggregation, the blocks are of equal lengths but, at each period, the oldest observation is dropped and the newest is included.

This paper focuses on the case of overlapping aggregation.

Although many papers examine continuous-valued time series, issues related to their application [11, 12] and different types of aggregation in them [2-9], the same is not true for time series of counts. Brännäs, Hellström and Nordström [13] first studied temporal and cross-sectional aggregation of an Integer Auto-Regressive process of order one, INAR(1). To our knowledge, forecast horizon aggregation in more general Integer Auto-Regressive Moving Average (INARMA) models has not been studied before. This is the motivation for this study, to begin to address this issue and present some new results. The paper is structured as follows. The forecast horizon aggregation of INARMA(p,q) processes is discussed in detail in section 2. A simulation experiment is designed and performed in section 3 to assess the accuracy of the aggregated forecasts of section 2. An empirical analysis, based on two datasets, is performed in section 4. The conclusions are provided in the final section of the paper.

2. FORECAST HORIZON AGGREGATION AND FORECASTING

In this section, aggregation and forecasting over a forecast horizon is discussed. This has applications in many areas. Some application areas require forecasts of the whole distribution ([14, 15]). However, other application areas need forecasts of the conditional mean ([13, 16]). This study concentrates on estimation of the conditional mean, while further research will focus on forecasting the whole distribution. This research examines whether there is any benefit to be leveraged from INARMA models in forecasting the conditional mean. It is shown that there is such a benefit in certain circumstances.

First, the conditional expected values of the aggregated INAR(1) and INMA(1) processes are presented. Then, it is shown that the forecast horizon aggregation of an INARMA process is an INARMA process. The conditional expected value of the aggregated INARMA process is also obtained.

The most common forecasting procedure discussed in the time series literature is using the conditional expectation. The main advantage of this method, apart from being simple, is that it produces forecasts with minimum mean square error (MMSE). This forecasting procedure is adopted in this paper.

2.1 INAR(1) Models

A Poisson INAR(1) process, PoINAR(1) is defined by:

$$\mathrm{Y_t} = \alpha \circ \mathrm{Y_{t\text{-}1}} + \mathrm{Z_t} \quad (1)$$

where $\alpha \in (0,1]$ and $\{Z_t\}$ is a sequence of i.i.d. non-negative integer-valued Poisson distributed random variables, with mean and finite variance λ. Z_t and Y_{t-1} are assumed to be stochastically independent for all points in time. The thinning operation "$\circ$" of Sueutel and van Harn [17] is defined by $\alpha \circ Y = \sum_{i=1}^{Y} X_i$ where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables with $P(X_i = 1) = \alpha$ for $i = 1, \dots, Y$.

It follows from [14] that the conditional mean of the aggregated process is:

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

As shown in Appendix A, the conditional variance of the aggregated process is as follows:

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

$$+ \frac{2\lambda}{1-\alpha} \sum_{j=1}^{l} \alpha^{2j-1} \left[(l-j) - \frac{\alpha(1-\alpha^{l-j})}{1-\alpha}\right] \quad (3)$$

For an INAR(1) process of (1), the cumulative value over horizon l is given by:

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

$$+ \cdots + (\alpha^l \circ Y_t + \alpha^{l-1} \circ Z_{t+1} + \alpha^{l-2} \circ Z_{t+2} + \cdots + Z_{t+l}) \quad (4)$$

Bearing in mind that $\alpha \circ X + \beta \circ X \neq (\alpha + \beta) \circ X$ (the LHS is the sum of two Binomial random variables with the same number of trials and different success probabilities), the above equation can be written in the following form:

$$\sum_{j=1}^{l} Y_{t+j} = \sum_{j=1}^{l} \sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l} \sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Z_{t+k_{ij}} \quad (5)$$

where n_j^1 is the number of Y_t terms in each of $\{Y_{t+j}\}_{j=1}^{l}$ in (4), 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 $\{Y_{t+j}\}_{j=1}^{l}$ in (4), and ψ_{ij}^2 is the corresponding coefficient for each $Z_{t+k_{ij}}$. Further details about the coefficients are given by [18].

It can be seen that, based on (5), the conditional expected value of the aggregated PoINAR(1) process is:

$$E\left(\sum_{j=1}^{l} Y_{t+j} \,|Y_t\right) = \left(\sum_{j=1}^{l} \sum_{i=1}^{n_j^1} \psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l} \sum_{i=1}^{n_j^2} \psi_{ij}^2\right) \lambda \quad (6)$$

The above equation is the same as (2). At time T, when Y_T is observed, the aggregated forecast can be obtained from:

$$E\left(\sum_{j=1}^{l} Y_{T+j} \,|Y_T\right) = \frac{\alpha(1-\alpha^l)}{1-\alpha} Y_T + \frac{\lambda}{1-\alpha}\left[l - \sum_{j=1}^{l} \alpha^j\right] \quad (7)$$

2.2 INMA(1) Models

For an INMA(1) process of $Y_t = \beta \circ Z_{t-1} + Z_t$, where $\beta \in (0,1]$ and $\{Z_t\}$ is as before, the cumulative value over horizon l is given by:

$$\sum_{j=1}^{l} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l} = (\beta \circ Z_t + Z_{t+1}) + (\beta \circ Z_{t+1} + Z_{t+2})$$

$$+ \cdots + (\beta \circ Z_{t+l-1} + Z_{t+l}) \quad (8)$$

The above equation can be written in the following form:

$$\sum_{j=1}^{l} Y_{t+j} = \sum_{j=1}^{l} \sum_{i=1}^{n_j} \psi_{ij} \circ Z_{t+k_{ij}} \quad (9)$$

where n_j is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ and is the corresponding coefficient for each $Z_{t+k_{ij}}$. Further details about the coefficients are given by [18].

Based on (9), the conditional expected value of the aggregated INMA(1) process is:

$$E\left(\sum_{j=1}^{l} Y_{t+j} \,|Y_t\right) = \left(\sum_{j=1}^{l} \sum_{i=1}^{2} \psi_{ij}\right)\lambda = \left(\sum_{j=1}^{l} (1+\beta)\right)\lambda = l(1+\beta)\lambda \quad (10)$$

2.3 INARMA(*p*,*q*) Models

This paper examines aggregation and forecasting of a general INARMA process over a forecast horizon. The INARMA(p,q) process is given by:

$$Y_t = \sum_{i=1}^{p} \alpha_i \circ Y_{t-i} + Z_t + \sum_{j=1}^{q} \beta_j \circ Z_{t-j} \quad (11)$$

where $\alpha_1, \ldots, \alpha_{p-1} \in [0,1], \alpha_p \in (0,1]; \beta_1, \ldots, \beta_{q-1} \in [0,1], \beta_q \in (0,1]$ 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. The thinning operations are defined as follows:

$$\alpha \circ Y = \sum_{i=1}^{Y} X_i \tag{12}$$

where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables with $P(X_i = 1) = \alpha$ for $i = 1, \dots, Y$. This paper follows the approach of Du and Li [19] and McKenzie [20] regarding the Binomial thinning mechanisms for INAR(p) and INMA(q), respectively. Therefore, it is assumed that the individual thinning operations $\alpha_i \circ Y_{t-i}$ for $i = 1, \dots, p$ and $\beta_j \circ Z_{t-j}$ for $j = 1, \dots, q$ are performed independently not only from each other, but also from corresponding operations at previous times in (11).

The stationarity conditions of this process are the same as those of an INAR(p) process. Neal and Rao [21] suggest that the invertibility conditions for this process are the same as the those of an MA(q) process ($\sum_{j=1}^{q} \beta_j < 1$).

The MMSE one-step-ahead forecast for an INARMA(p,q) process of (11) is:

$$\hat{Y}_{T+1} = \alpha_1 Y_T + \cdots + \alpha_p Y_{T-p+1} + \lambda + \beta_1 Z_T + \cdots + \beta_q Z_{T-q+1} \tag{13}$$

The h-step ahead forecast when $h \le q$ is:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \cdots + \alpha_p Y_{T+h-p} + \lambda + \beta_h Z_T + \cdots + \beta_q Z_{T+h-q} + \lambda(\beta_1 + \cdots + \beta_{h-1}) \tag{14}$$

where the Y values on the RHS of (14) may be either actual or forecast values. When $h > q$, the h-step ahead forecast becomes:

$$\hat{Y}_{T+h} = \alpha_1 Y_{T+h-1} + \cdots + \alpha_p Y_{T+h-p} + \lambda \sum_{j=0}^{q} \beta_j \tag{15}$$

where again the Y values on the RHS of the above equation may be either actual or forecast values and $\beta_0 = 1$.

We next present two propositions regarding the aggregation and forecasting of an

INARMA(p,q) process.

Proposition 1. Aggregation of an INARMA(p,q) process over a forecast horizon results in an INARMA(p,q) process.

Proof.

For an INARMA(p,q) process of (11), the aggregated process over a forecast horizon can be written as:

$$\sum_{j=1}^{l} Y_{t+j} = \sum_{j=1}^{l} \left\{ \left[\sum_{i=1}^{p} \alpha_i \circ Y_{t+j-i} \right] + Z_{t+j} + \left[\sum_{i=1}^{q} \beta_i \circ Z_{t+j-i} \right] \right\} =$$

$$\sum_{i=1}^{p} \left[\alpha_i \circ \sum_{j=1}^{l} Y_{t+j-i} \right] + \sum_{j=1}^{l} Z_{t+j} + \sum_{i=1}^{q} \left[\beta_i \circ \sum_{j=1}^{l} Z_{t+j-i} \right] \qquad (16)$$

Now, if we assume that $\sum_{j=1}^{l} Y_{t+j} = Y_\tau$ and $\sum_{j=1}^{l} Z_{t+j} = Z_\tau$, (16) can be written as:

$$Y_\tau = \sum_{i=1}^{p} \alpha_i \circ Y_{\tau-i} + Z_\tau + \sum_{i=1}^{q} \beta_i \circ Z_{\tau-i} \qquad (17)$$

which is also an INARMA(p,q) process. Therefore, aggregation of an INARMA(p,q) process over a forecast horizon results in an INARMA(p,q) process with the same INAR and INMA parameters but with a different innovation parameter. If $Z_t \sim Po(\lambda)$, Z_τ will be the sum of l independent Poisson variables; thus, $Z_\tau \sim Po(l\lambda)$.

Proposition 2. The forecast horizon aggregated INARMA(p,q) process can be written in terms of the last p observations as follows:

$$\sum_{j=1}^{l} Y_{t+j} = \sum_{j=1}^{l} \sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l} \sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Y_{t-1} + \cdots$$

$$+ \sum_{j=1}^{l} \sum_{i=1}^{n_j^p} \psi_{ij}^p \circ Y_{t-p+1} + \sum_{j=1}^{l} \sum_{i=1}^{n_j^{p+1}} \psi_{ij}^{p+1} \circ Z_{t+k_{ij}} \qquad (18)$$

with the parameters as shown in Table 1 (see Appendix B for the proof).

The conditional expected value of the aggregated process given the *p*-previous observations can then be obtained from:

$$E\left(\sum_{j=1}^{l} Y_{t+j}\,|Y_{t-p+1},\dots,Y_{t-1},Y_t\right)=\left(\sum_{j=1}^{l}\sum_{i=1}^{n_j^1}\psi_{ij}^1\right)Y_t+\left(\sum_{j=1}^{l}\sum_{i=1}^{n_j^2}\psi_{ij}^2\right)Y_{t-1}+\cdots$$

$$+\left(\sum_{j=1}^{l}\sum_{i=1}^{n_j^p}\psi_{ij}^p\right)Y_{t-p+1}+\left(\sum_{j=1}^{l}\sum_{i=1}^{n_j^{p+1}}\psi_{ij}^{p+1}\right)\lambda \qquad (19)$$

The above equation is then used to forecast the aggregated process. In the next section, the accuracy of such aggregated forecasts will be assessed for INAR(1), INMA(1) and INARMA(1,1) processes.

Table 1 Parameters of the forecast horizon aggregated INARMA(*p*,*q*) model

for $w=1,\dots,p$	$n_j^w=\begin{cases}\left(\sum_{i=1}^{p} n_{j-i}^w\right)+1 & j\le p-(w-1)\\ \sum_{i=1}^{p} n_{j-i}^w & j>p-(w-1)\end{cases}$	$\psi_{ij}^w=\begin{cases}\begin{bmatrix}\alpha_p\psi_{i(j-p)}^w & i=1,\dots,n_{j-p}^w\\ \vdots & \vdots\\ \alpha_1\psi_{i(j-1)}^w & i=n_{j-2}^w+1,\dots,n_{j-2}^w+n_{j-1}^w\\ \alpha_{j+(w-1)} & i=n_{j-1}^w+1\end{bmatrix} & j\le p-(w-1)\\ \begin{bmatrix}\alpha_p\psi_{i(j-p)}^w & i=1,\dots,n_{j-p}^w\\ \vdots & \vdots\\ \alpha_1\psi_{i(j-1)}^w & i=n_{j-2}^w+1,\dots,n_{j-2}^w+n_{j-1}^w\end{bmatrix} & j>p-(w-1)\end{cases}$
	$n_j^{p+1}=\left(\sum_{i=1}^{p} n_{j-i}^{p+1}\right)+(q+1)$	$\psi_{ij}^{p+1}=\begin{cases}\alpha_p\psi_{i(j-p)}^{p+1} & i=1,\dots,n_{j-p}^{p+1}\\ \vdots & \vdots\\ \alpha_1\psi_{i(j-1)}^{p+1} & i=n_{j-2}^{p+1}+1,\dots,n_{j-2}^{p+1}+n_{j-1}^{p+1}\\ \beta_q,\dots,\beta_1,1 & i=n_{j-1}^{p+1}+1,\dots,n_{j-1}^{p+1}+n_j^{p+1}\end{cases}$ $k_{ij}=\begin{cases}\{k_{i(j-p)}\} & i=1,\dots,n_{j-p}^{p+1}\\ \vdots & \vdots\\ \{k_{i(j-1)}\} & i=\sum_{z=2}^{p} n_{j-z}^{p+1}+1,\dots,(\sum_{z=2}^{p} n_{j-z}^{p+1})+n_{j-1}^{p+1}\\ j-q,\dots,j-1,j & i=\sum_{z=1}^{p} n_{j-z}^{p+1}+1,\dots,n_j^{p+1}\end{cases}$

3. SIMULATION

In order to test the benefit of using the INARMA horizon-aggregated forecasts for short histories as well as long, Monte Carlo simulations are conducted. The results for the INARMA horizon-aggregated forecasts (hereafter abbreviated as INARMA-Agg), have been compared to the results of using cumulative h-step ahead forecasts over the horizon (hereafter abbreviated as INARMA-h). The latter is given by $\sum_{i=1}^{l} \hat{Y}_{t+i}$, where $\hat{Y}_{t+i}$ is the i-step ahead forecast. It should be emphasized that there are two approaches in the literature regarding h-step ahead forecasting. The first approach used by Brännäs and Hellström [22], is based on repeated substitution of the INARMA process. For example, the h-step ahead forecast of an INAR(1) process can be obtained from:

$$Y_{T+h} = \alpha^{h} \circ Y_T + \sum_{i=1}^{h} \alpha^{h-i} \circ Z_{T+i} \qquad (20)$$

It can be easily shown that aggregation of (20) over a horizon results in (7). Therefore, this approach results in the same aggregated forecast as that proposed by this study.

In the second approach, the Y value on the right hand side of the equation $\hat{Y}_{T+h} = \alpha \hat{Y}_{T+h-1} + \lambda$, is a forecast [19, 23]. The h-step ahead forecasts in this section are calculated based on this approach. For simulation purposes, it is assumed that the innovations, $\{Z_t\}$, have a Poisson distribution with parameter λ. Although this assumption is restrictive, a large number of data series of the empirical datasets used in this study met the above condition (see section 4). This is consistent with the larger empirical study by Eaves [24] discussed in section 4. The theoretical findings in this paper, however, are not based on any distributional assumptions and can be used as a framework for future studies based on other marginal distributions.

Three INARMA process are considered in the simulation experiment: INAR(1), INMA(1) and INARMA(1,1). The aggregated forecasts for these models can be derived from section 2. For an INAR(1) process the aggregated forecast is provided by (2). For the INMA(1) and INARMA(1,1) processes, these are given by:

$$E\left[\sum_{i=1}^{l} Y_{T+i} \,|Y_T\right] = l(1+\beta)\lambda \quad (21)$$

$$E\left[\sum_{i=1}^{l} Y_{T+i} \,|Y_T\right] = \frac{\alpha(1-\alpha^l)}{1-\alpha} Y_T + \frac{\lambda(1+\beta)}{1-\alpha}\left[l - \sum_{j=1}^{l} \alpha^j\right] \quad (22)$$

The control parameters to be varied are α, β, λ and n (number of historical observations). The range of these parameters is given in Table 2. Different lengths of series are considered in the Monte Carlo simulations to test the sensitivity of the results to the length of history. In real cases, we are often restricted by short lengths of history (as will be seen in the empirical analysis). Hence, we use $n = 24, 36, 48, 96, 500$ to encompass short data histories as well as long. The forecast horizons considered are three, six and nine periods. The number of replications is set to 1000. This is consistent with other studies of INARMA processes which used the same or fewer replications (eg. [22, 25-27]) and have been found to give reliable results when compared with findings known from theory.

The data series are divided into two periods: estimation period and performance period. Initialization and estimation of parameters are conducted in the estimation period and the forecasting accuracy is assessed in the performance period. If at least two non-zero values 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 values are observed in the estimation period, this period will be extended until the second non-zero value is observed.

As an example consider the case of $n = 24$ and $l = 3$. Under this experimental scenario the length of the estimation period is 12 (if at least two non-zero values are observed, else it is extended until two such values are available). The forecast errors are then calculated in the performance block of periods from period $t = 13$ (for periods 13, 14 and 15) to period $t = 22$ (for periods 22, 23 and 24).

Table 2 The range of control parameters

Number of observations	$n = 24, 36, 48, 96, 500$			
INAR(1)	$\alpha = 0.1, \lambda = 0.5$ $\alpha = 0.5, \lambda = 0.5$	$\alpha = 0.1, \lambda = 1$ $\alpha = 0.5, \lambda = 1$	$\alpha = 0.1, \lambda = 3$ $\alpha = 0.5, \lambda = 3$	$\alpha = 0.1, \lambda = 5$ $\alpha = 0.5, \lambda = 5$
INMA(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)	$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$ $\alpha = 0.1, \beta = 0.9, \lambda = 0.5$ $\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	$\alpha = 0.1, \beta = 0.1, \lambda = 1$ $\alpha = 0.1, \beta = 0.9, \lambda = 1$ $\alpha = 0.5, \beta = 0.5, \lambda = 1$	$\alpha = 0.1, \beta = 0.1, \lambda = 5$ $\alpha = 0.1, \beta = 0.9, \lambda = 5$ $\alpha = 0.5, \beta = 0.5, \lambda = 5$	

All coding is in MATLAB and random numbers are generated by the *poissrnd* function for the Poisson distribution and *binornd* function for the Binomial thinning operations.

The autoregressive, moving average and innovation parameters are estimated using the Yule-Walker (YW) estimation method and the estimates are updated in each period. This is a simple estimation method and it has been shown that the YW estimators are asymptotically equivalent to the Conditional Least Squares (CLS) estimators for an INAR(1) process [28]. Also, for INMA(1) and INARMA(1,1) processes, the forecasts produced by the two methods are close in terms of Mean Square Error [18].

The forecast accuracy of the two aggregated forecasting methods (INARMA-Agg and INARMA-h) is compared in terms of Mean Square Error (MSE). MSE is a widely used measure in the forecasting literature, is mathematically easy to handle and is a sensible measure for evaluating an individual time series. For methods that produce unbiased forecasts, the MSE coincides with the forecast error variance. In the empirical analysis presented in section 4, the empirical bias properties of the INARMA methods are checked.

Table 3 compares the MSE of the two methods for an INAR(1) process when $l = 3$. The results for the cases where $l = 6, 9$ are presented in Appendix C.

Table 3 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INAR(1) series when $l = 3$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.9929	0.9937	0.9950	0.9938	0.9960
$\alpha = 0.5, \lambda = 0.5$	0.8398	0.8527	0.8367	0.8201	0.7885
$\alpha = 0.1, \lambda = 1$	0.9785	0.9887	0.9924	0.9935	0.9921
$\alpha = 0.5, \lambda = 1$	0.8278	0.7953	0.7906	0.7410	0.7216
$\alpha = 0.1, \lambda = 3$	0.9631	0.9604	0.9727	0.9757	0.9835
$\alpha = 0.5, \lambda = 3$	0.7015	0.6536	0.6328	0.5862	0.5395
$\alpha = 0.1, \lambda = 5$	0.9231	0.9311	0.9473	0.9614	0.9734
$\alpha = 0.5, \lambda = 5$	0.5954	0.5512	0.5184	0.4768	0.4397

The results of Table 3 show that for an INAR(1) process, the INARMA-Agg method outperforms the INARMA-h method for all parameter ranges. The improvement increases with the value of λ and the length of history. Also, the improvement is higher when the autoregressive parameter is higher because of the nonlinearity of the model. For large values of h, the INARMA-h forecasts converge to the unconditional mean of the INAR(1) process:

$$\hat{Y}_{T+h} \to \frac{\lambda}{1-\alpha}$$

The larger the value of α, the more variable the data, which makes this convergence less desirable. Some authors have suggested using different models for different horizons to improve forecast accuracy [29-31].

Next, the INARMA-Agg and INARMA-h methods are compared for an INMA(1) process when $l = 3$ (See Appendix C for the cases of $l = 6, 9$). It can be seen from Table 4 that for an INMA(1) process, INARMA-Agg method has very slightly better forecasts in terms of MSE than the INARMA-h method. The former is based on (21) and the latter has the following expression:

$$\sum_{i=1}^{l} \hat{Y}_{T+i} = \sum_{i=1}^{l} \hat{\beta}_{T+i} \hat{Z}_{T+i} + \hat{\lambda}_{T+i} \tag{23}$$

Comparing (21) and (23) reveals that the only difference between the two forecasting methods is that the INARMA-h method uses the forecast of the innovation term ($\hat{Z}_{T+i}$), while the INARMA-Agg method uses $\hat{\lambda}$ as an estimate for the innovation term. This difference remains for large samples, because $\hat{\lambda}$ will converge to a constant (the true λ), but $\hat{Z}_{T+1}$, say, would remain a random variable. However, it is expected that the two methods should be very close and the results of Table 4 confirm this.

Table 4 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INMA(1) series when $l = 3$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.9995	0.9996	0.9998	0.9999	1.0000
$\beta = 0.5, \lambda = 0.5$	0.9969	0.9981	0.9985	0.9991	0.9998
$\beta = 0.9, \lambda = 0.5$	0.9944	0.9961	0.9968	0.9982	0.9996
$\beta = 0.1, \lambda = 1$	0.9992	0.9997	0.9998	0.9999	1.0000
$\beta = 0.5, \lambda = 1$	0.9971	0.9979	0.9982	0.9991	0.9998
$\beta = 0.9, \lambda = 1$	0.9937	0.9958	0.9967	0.9982	0.9996
$\beta = 0.1, \lambda = 3$	0.9988	0.9994	0.9998	0.9999	1.0000
$\beta = 0.5, \lambda = 3$	0.9961	0.9981	0.9982	0.9991	0.9998
$\beta = 0.9, \lambda = 3$	0.9932	0.9947	0.9966	0.9981	0.9996
$\beta = 0.1, \lambda = 5$	0.9996	0.9993	0.9997	0.9999	1.0000
$\beta = 0.5, \lambda = 5$	0.9953	0.9975	0.9984	0.9989	0.9998
$\beta = 0.9, \lambda = 5$	0.9929	0.9950	0.9964	0.9979	0.9996

It can also be seen from Table 4 that with an increase in β, the MSE of the INARMA-Agg method slightly improves compared to that of an INARMA-h method. This could also be attributed to the fact that, for large values of h, the INARMA-h forecasts converge to the unconditional mean of the INMA(1) process $\lambda(1 + \beta)$. Again, larger values of β produce more variable data; therefore, this convergence would result in less accurate forecasts.

Finally, Table 5 compares the MSE of INARMA-Agg and INARMA-h methods for an INARMA(1,1) process when $l = 3$. The results confirm the above arguments that when the autoregressive parameter is high, the INARMA-Agg method has smaller MSE than the INARMA-h method and the improvement increases with the length of horizon. However, for small autoregressive and moving average parameters, the INARMA-h method is better than the INARMA-Agg. Again, when the number of observation increases the difference decreases.

Table 5 $\mathrm{MSE}_{Agg}/\mathrm{MSE}_h$ of aggregated forecasts for INARMA(1,1) series when $l = 3$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	1.1288	1.0764	1.0565	1.0159	0.9976
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	1.0045	0.9895	0.9661	0.9551	0.9718
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.8688	0.8342	0.8209	0.8145	0.8129
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0946	1.0590	1.0474	1.0144	0.9969
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	1.0402	0.9692	0.9641	0.9611	0.9735
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8646	0.8623	0.8402	0.8285	0.8134
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	1.0425	1.0258	1.0292	1.0042	0.9959
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9790	0.9680	0.9514	0.9561	0.9639
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8652	0.8525	0.8481	0.8252	0.8013

The above results along with the results of Appendix C suggest that for an INAR(1) process, the INARMA-Agg method outperforms the INARMA-h method in terms of MSE. The difference between two methods is high when the autoregressive parameter is high. However, for an INMA(1) process, the two methods produce very close forecasts. With an increase in the moving average parameter the improvement of the INARMA-Agg method over the INARMA-h method slightly increases.

For an INARMA(1,1) process, when the length of horizon is short and the autoregressive and moving average parameters are small, the INARMA-h forecasts have smaller MSEs than INARMA-Agg forecasts. For all the other cases, the latter method beats the former method using MSE.

4. EMPIRICAL ANALYSIS

In this section, an empirical analysis is conducted to validate the findings on real data. The real demand data series for this research consists of the Royal Air Force (RAF) individual demand histories of 16,000 Stock Keeping Units (SKUs) over a period of 6 years (monthly observations). We have also used another dataset which consists of 3,000 real intermittent demand data series from the automotive industry[1] (from [32]) 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 datasets are called Dataset 1 and Dataset 2 from now on.

As previously mentioned, this paper 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 INARMA processes of this study (INARMA(0,0), INAR(1), INMA(1), and INARMA(1,1)), three of them have a Poisson distribution when the innovation terms are Poisson. The only exception is the INARMA(1,1) process where:

$$\frac{\mathrm{var}(Y_t)}{E(Y_t)} = \frac{1+\alpha+\beta+3\alpha\beta}{1+\alpha+\beta+\alpha\beta} \leq 1.5 \qquad \text{for } 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1 \tag{24}$$

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 distributed, the test statistic:

[1] This dataset is available from: http://www.forecasters.org/ijf/data/Empirical%20Data.xls

$$T_{CC} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\bar{X}} \tag{25}$$

has a chi-square distribution with $(n-1)$ degrees of freedom. Therefore, H_0 is rejected if $T_{CC} > \chi^2_{n-1;1-A}$.

where A is the significance level. 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:

$$T_{CCR} = \frac{T_{CC}}{1.5} \tag{26}$$

The new statistic also has a chi-square distribution with $(n-1)$ degrees of freedom.

The above filtering, with $A = 0.05$, results in the exclusion of some data series. Out of the 16,000 series, 12,800 series remained and out the 3,000 series, 1,943 series remained. As mentioned in section 3, a high percentage of series used in this study can be modelled using the Poisson assumption. This is consistent with the study by Eaves [24], in which over 80% of series had lead-time demand fitting the Poisson distribution at the 5% significance level.

Further filtering of data was performed for series with fewer than two nonzero demands. 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. Relevant characteristics of the filtered datasets are summarized in Table 6.

Table 6 Information about filtered 16,000 and 3,000 datasets

	Min	Mean	Max	Percentage of zeros
Dataset 1: 16,000 series (5,168 filtered series)	0	0.2177	14	85.62

Dataset 2: 3,000 series (1,943 filtered series)	0	2.0194	2	22.77

The design of the empirical analysis follows the detailed simulation design of section 3. The Yule-Walker estimation method has been used to estimate the parameters of the INARMA models. Two values for forecast horizon have been considered: $l = 3, 6$.

The appropriate INARMA model needs to be identified among the four possible candidates. This is done using a two-stage identification procedure [18]. The first stage distinguishes between the INARMA(0,0) and the other INARMA models. The Ljung-Box statistic of:

$$Q^* = n(n+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{n-j} \qquad (27)$$

is used for this reason. This is a standard test used for conventional ARMA models that is included in most software packages (including MATLAB which is used in this paper) and, based on the argument by Latour [33], it can be used for INARMA models as well. The AIC, as calculated by the formula $\mathrm{AIC} \approx N\log\hat{\sigma}_a^2 + 2m$ is then used for identification among the other INARMA models. This is again based on the argument of Latour [33] 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. [16]).

This identification procedure is applied on our empirical data and the results in terms of the percentage of each of four INARMA models for each dataset are presented in Table 7.

Table 7 Identification results* for Dataset 1 and Dataset 2

Models	INARMA(0,0)	INAR(1)	INMA(1)	INARMA(1,1)
Dataset 1	98.12	0.66	1.04	0.17
Dataset 2	54.55	23.88	17.96	3.60

*in terms of percentage of series

As shown in Table 7, the great majority of series in Dataset 1 are identified as INARMA(0,0) which was expected due to high number of zeros. The INARMA-Agg and INARMA-h methods produce the same result for INARMA(0,0) series; therefore, these methods are only compared for the other three INARMA models (INAR(1), INMA(1) and INARMA(1,1)). The MSE results are compared in Table 8 and Table 9 for $l = 3$ and $l = 6$, respectively. The bias, in terms of Mean Error, has been checked and found to be low (see [18]).

Table 8 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INARMA series when $l = 3$

Models	INAR(1)	INMA(1)	INARMA(1,1)
Dataset 1	0.9636	0.9907	1.5557
Dataset 2	0.9706	0.9841	1.1180

Table 9 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INARMA series when $l = 6$

Models	INAR(1)	INMA(1)	INARMA(1,1)
Dataset 1	0.9057	0.9789	1.6815
Dataset 2	0.9242	0.9573	1.2176

In order to compare the results with simulation results, the range of estimated parameters for each of the INARMA models are provided in Table 10.

Table 10 Parameters' estimates for Dataset 1 and Dataset 2

Models	INAR(1)	INMA(1)	INARMA(1,1)
Dataset 1	$\hat{\alpha}$ is close to 0.2 (the average is 0.2460 and 52.94 percent are between 0.1 and 0.3)	$\hat{\beta}$ is close to zero (the average is 0.0898 and 46.29 percent are between 0 and 0.1)	$0.1 < \hat{\alpha} < 0.3$ (the average is 0.2988 and 66.67 percent are between 0.05 and 0.35)
	$\hat{\lambda}$ is around 0.5 (the average is 0.3562 and 97.06 percent are between 0 and 1)	$\hat{\lambda}$ is around 0.3 (the average is 0.3782 and 55.56 percent are between 0.2 and 0.4)	$\hat{\beta}$ is close to zero (the average is 0.1405 and 77.78 percent are between 0 and 0.1)
			$\hat{\lambda}$ is around 0.3 (the average is 0.3558 and 44.44 percent are between 0.2 and 0.5)
Dataset 2	$\hat{\alpha}$ is close to 0.1 (the average is 0.1234 and 50.65 percent are between 0.05 and 0.15)	$\hat{\beta}$ is close to zero (the average is 0.0374 and 79.94 percent are between 0 and 0.05)	$0.1 < \hat{\alpha} < 0.3$ (the average is 0.1907 and 54.28 percent are between 0.05 and 0.35)

	$\hat{\lambda}$ is around 2 (the average is 2.5972 and 40.52 percent are between 1 and 3)	$\hat{\lambda}$ is between 2 and 3 (the average is 2.7357 and 43.55 percent are between 2 and 3)	$\hat{\beta}$ is close to zero (the average is 0.0773 and 57.14 percent are between 0.01 and 0.1)
			$\hat{\lambda}$ is around 2 (the average is 2.1996 and 67.14 percent are between 1 and 2.5)

It can be seen that the results for the INAR(1) and INMA(1) processes are in agreement with the simulation results. For the INAR(1) series, the results are comparable to the simulation results of Table 3 and Table C-1. For the INMA(1) series, this is comparable to the simulation results of Table 4 and Table C-3. Finally, for the INARMA(1,1) series the results of Dataset 2 are comparable to the simulation results of Table 5 and Table C-5 which suggest that INARMA-h produces better results than INARMA-Agg for those parameter ranges. It is worth mentioning that for Dataset 1, only a few series were identified as INARMA(1,1).

Therefore, the above results suggest that for an INAR(1) process, the INARMA-Agg method has lower MSE that the INARMA-h method and the improvement increases with the length of history. For an INMA(1) process, the two methods are very close. For INARMA(1,1) series, the empirical results of Dataset 2 confirm the simulation result that, for low autoregressive and moving average parameters, INARMA-h outperforms INARMA-Agg in terms of MSE.

5. CONCLUSIONS

This paper addresses forecast horizon aggregation in INARMA processes. The conditional mean of the aggregated process is obtained for the general INARMA(p,q) process which can be used for forecasting. The purpose of the paper is not to propose new forecasting methods, but rather to compare the performance of alternative INARMA approaches. This has been achieved by simulation and empirical analysis.

It is shown that the aggregation of an INARMA process over a horizon results in an INARMA process. The conditional mean of the aggregated process is also derived as a basis for forecasting. The results of a simulation experiment are provided to assess the accuracy of the forecasts produced using the conditional mean of the aggregated process for three INARMA processes: INAR(1), INMA(1) and INARMA(1,1). The results are compared to the case where the forecasts are produced by adding up the *h*-step-ahead forecasts over the forecast horizon.

The simulation results suggest that, in most cases, the aggregation method generates forecasts with smaller MSEs than the cumulative *h*-step-ahead method. The difference is substantial when the autoregressive parameter is high. The only case in which the INARMA-h method is better than the INARMA-Agg method is for an INARMA(1,1) process with small autoregressive and moving average parameters and short length of forecast horizon.

The performance of these forecasts is also tested on empirical data of two real demand data series and the results generally confirm the simulation results.

As previously mentioned, this paper has focused on INARMA processes with Poisson innovations. Other discrete self-decomposable distributions such as generalized Poisson and negative binomial distributions could be used as marginal distributions. Also, the findings of this paper are based on MSE. Other performance measures could be used to examine the accuracy of forecasts. In an inventory management context, this can be done by looking at inventory implication metrics such as service level and inventory level [34].

Finally, aggregated forecasts are becoming increasingly important for Decision Support Systems (DSS) in the area of production planning [35]. Further research into issues related to the application of aggregated forecasts in such a context should be very important both from academic and practitioner perspectives.

Acknowledgements

The authors would like to thank the three anonymous referees for their comments, which have assisted in sharpening the statements of motivation, scope and contribution of the paper.

APPENDIX A- HORIZON FORECASTING FOR AN INAR(1) MODEL

In this appendix, it is shown how to derive the conditional second moment of a forecast horizon aggregated PoINAR(1) process ($\mu_Z = \sigma_Z^2 = \lambda$). The aggregated process over horizon l can be written as:

$$\sum_{i=1}^{l} Y_{t+i} \overset{d}{=} (\alpha \circ Y_t + Z_{t+1}) + (\alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + Z_{t+2}) + \cdots$$
$$+(\alpha^l \circ Y_t + \alpha^{l-1} \circ Z_{t+1} + \alpha^{l-2} \circ Z_{t+2} + \cdots + \alpha \circ Z_{t+l-1} + Z_{t+l}) \qquad \text{(A.1)}$$

where $\overset{d}{=}$ means equal in distribution. It can be simplified as:

$$\sum_{i=1}^{l} Y_{t+i} \overset{d}{=} (\alpha \circ Y_t + \alpha^2 \circ Y_t + \cdots + \alpha^l \circ Y_t)$$
$$+(Z_{t+1} + \alpha \circ Z_{t+1} + \cdots + \alpha^{l-1} \circ Z_{t+1}) + (Z_{t+2} + \alpha \circ Z_{t+2} + \cdots + \alpha^{l-2} \circ Z_{t+2}) +$$
$$\cdots + (Z_{t+l-1} + \alpha \circ Z_{t+l-1}) + Z_{t+l} \qquad \text{(A.2)}$$

We know that $\text{cov}(\alpha^i \circ X, \alpha^j \circ X) = \alpha^i \alpha^j E(X^2) - \alpha^i E(X) \alpha^j E(X) = \alpha^i \alpha^j \text{var}(X)$. Hence, we have $\text{cov}(\alpha^i \circ Z_t, \alpha^j \circ Z_t) = \alpha^i \alpha^j \lambda$. The variance of the (A.2) given Y_t is:

$$\text{var}\left[\sum_{i=1}^{l} Y_{t+i} \,|Y_t\right] = \text{var}(\alpha \circ Y_t|Y_t) + \text{var}(\alpha^2 \circ Y_t|Y_t) + \cdots + \text{var}(\alpha^l \circ Y_t|Y_t)$$
$$+2\text{cov}(\alpha \circ Y_t, \alpha^2 \circ Y_t|Y_t) + 2\text{cov}(\alpha \circ Y_t, \alpha^3 \circ Y_t|Y_t) + \cdots + 2\text{cov}(\alpha \circ Y_t, \alpha^l \circ Y_t|Y_t)$$
$$+ 2\text{cov}(\alpha^2 \circ Y_t, \alpha^3 \circ Y_t|Y_t) + 2\text{cov}(\alpha^2 \circ Y_t, \alpha^4 \circ Y_t|Y_t) + \cdots + 2\text{cov}(\alpha^2 \circ Y_t, \alpha^l \circ Y_t|Y_t)$$
$$+ \cdots + 2\text{cov}(\alpha^{l-2} \circ Y_t, \alpha^{l-1} \circ Y_t|Y_t) + 2\text{cov}(\alpha^{l-2} \circ Y_t, \alpha^l \circ Y_t|Y_t)$$
$$+2\text{cov}(\alpha^{l-1} \circ Y_t, \alpha^l \circ Y_t|Y_t)$$
$$+\text{var}(Z_{t+1}) + \text{var}(\alpha \circ Z_{t+1}) + \text{var}(\alpha^2 \circ Z_{t+1}) + \cdots + \text{var}(\alpha^l \circ Z_{t+1})$$
$$+2\text{cov}(Z_{t+1}, \alpha \circ Z_{t+1}) + 2\text{cov}(Z_{t+1}, \alpha^2 \circ Z_{t+1}) + \cdots + 2\text{cov}(Z_{t+1}, \alpha^l \circ Z_{t+1})$$

$$+2\text{cov}(\alpha \circ Z_{t+1}, \alpha^2 \circ Z_{t+1}) + 2\text{cov}(\alpha \circ Z_{t+1}, \alpha^3 \circ Z_{t+1}) + \cdots + 2\text{cov}\left(\alpha \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right)$$
$$+ \cdots + 2\text{cov}\left(\alpha^{l-2} \circ Z_{t+1}, \alpha^{l-1} \circ Z_{t+1}\right) + 2\text{cov}\left(\alpha^{l-2} \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right)$$
$$+2\text{cov}\left(\alpha^{l-1} \circ Z_{t+1}, \alpha^l \circ Z_{t+1}\right)$$

$$+\text{var}(Z_{t+2}) + \text{var}(\alpha \circ Z_{t+2}) + \cdots + \text{var}\left(\alpha^{l-1} \circ Z_{t+2}\right)$$
$$+2\text{cov}(Z_{t+2}, \alpha \circ Z_{t+2}) + 2\text{cov}(Z_{t+2}, \alpha^2 \circ Z_{t+2}) + \cdots + 2\text{cov}\left(Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right)$$
$$+2\text{cov}(\alpha \circ Z_{t+2}, \alpha^2 \circ Z_{t+2}) + 2\text{cov}(\alpha \circ Z_{t+2}, \alpha^3 \circ Z_{t+2}) + \cdots$$
$$+ 2\text{cov}\left(\alpha \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right)$$
$$+ \cdots + 2\text{cov}\left(\alpha^{l-3} \circ Z_{t+2}, \alpha^{l-2} \circ Z_{t+2}\right) + 2\text{cov}\left(\alpha^{l-3} \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right)$$
$$+2\text{cov}\left(\alpha^{l-2} \circ Z_{t+2}, \alpha^{l-1} \circ Z_{t+2}\right) + \cdots$$

$$+\text{var}(Z_{t+l-2}) + \text{var}(\alpha \circ Z_{t+l-2}) + \text{var}(\alpha^2 \circ Z_{t+l-2})$$
$$+2\text{cov}(Z_{t+l-2}, \alpha \circ Z_{t+l-2}) + 2\text{cov}(Z_{t+l-2}, \alpha^2 \circ Z_{t+l-2}) + 2\text{cov}(\alpha \circ Z_{t+l-2}, \alpha^2 \circ Z_{t+l-2})$$

$$+\text{var}(Z_{t+l-1}) + \text{var}(\alpha \circ Z_{t+l-1})$$
$$+2\text{cov}(Z_{t+l-1}, \alpha \circ Z_{t+l-1})$$
$$+\text{var}(Z_{t+l}) \qquad \text{(A.3)}$$

Since Y_t is fixed, $\text{cov}\left(\alpha^i \circ Y_t, \alpha^j \circ Y_t | Y_t\right) = \alpha^i \alpha^j \text{var}(Y_t | Y_t) = 0$. Hence:

$$\text{var}\left[\sum_{i=1}^{l} Y_{t+i} \,|Y_t\right] = \alpha(1-\alpha)E(Y_t|Y_t) + \alpha^2(1-\alpha^2)E(Y_t|Y_t) + \cdots + \alpha^l(1-\alpha^l)E(Y_t|Y_t)$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda] + \cdots + \left[\alpha^{2l}\lambda + \alpha^l\left(1-\alpha^l\right)\lambda\right]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^l]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^{l+1}]\lambda + \cdots + 2[\alpha^{2l-3} + \alpha^{2l-2}]\lambda$$
$$+ 2[\alpha^{2l-1}]\lambda$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda] + \cdots + \left[\alpha^{2l-2}\lambda + \alpha^{l-1}\left(1-\alpha^{l-1}\right)\lambda\right]$$
$$+2[\alpha + \alpha^2 + \cdots + \alpha^{l-1}]\lambda + 2[\alpha^3 + \alpha^4 + \cdots + \alpha^l]\lambda + \cdots + 2[\alpha^{2l-5} + \alpha^{2l-4}]\lambda$$
$$+ 2[\alpha^{2l-3}]\lambda + \cdots$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda] + [\alpha^4\lambda + \alpha^2(1-\alpha^2)\lambda]$$
$$+2[\alpha + \alpha^2]\lambda + 2[\alpha^3]\lambda$$
$$+\lambda + [\alpha^2\lambda + \alpha(1-\alpha)\lambda]$$
$$+2\alpha\lambda$$
$$+\lambda \qquad \text{(A.4)}$$

The above result can be summarized to:

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

$$+ \frac{2\lambda}{1-\alpha} \sum_{j=1}^{l} \alpha^{2j-1} \left[(l-j) - \frac{\alpha(1-\alpha^{l-j})}{1-\alpha}\right] \qquad \text{(A.5)}$$

APPENDIX B- HORIZON FORECASTING FOR AN INARMA(p,q) MODEL

In order to find the conditional mean of the forecast horizon 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:

$$\sum_{j=1}^{l} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l} \qquad \text{(B.1)}$$

To write the above equation in the form of (19), we need to know:

- the number and the coefficient of $\{Y_{t-w+1}\}_{w=1}^{p}$, and
- the number, the coefficient and the subscript of $Z_{t+k_{ij}}$

in (B.1). Each of these is discussed in the following subsections.

B.1 The Number of $\{Y_{t-w+1}\}_{w=1}^{p}$

Each of the $\{Y_{t+j}\}_{j=1}^{l}$ in the RHS of (B.1) needs to be expressed in terms of $\{Y_{t-w+1}\}_{w=1}^{p}$ by repeated substitution of Y_{t+j} in the equation for the INARMA(p,q) model (11). Because the autoregressive order of the process is p, Y_{t+j} can be expressed in terms of p previous observations by utilising the first component of the RHS of (11), namely: $\alpha_1 \circ Y_{t+j-1} + \cdots + \alpha_p \circ Y_{t+j-p}$.

Now, if $j \leq p - (w-1)$, there is one $Y_{t-(w-1)}$ when we express the jth observation in the RHS of (B.1) (Y_{t+j}) without any need for further 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, when $j > p - (w-1)$, each Y_{t+j} from (B.1) should be substituted by (11) in order to reach $Y_{t-(w-1)}$, and the number of $Y_{t-(w-1)}$ in each of $\{Y_{t+j}\}_{j=1}^{l}$ would be equal to the number of $Y_{t-(w-1)}$ in its p previous observations.

B.2 The Coefficient of $\{Y_{t-w+1}\}_{w=1}^{p}$

For $j \leq p - (w-1)$, the corresponding coefficient of $Y_{t-(w-1)}$ in the jth observation in the RHS of (B.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 $\{Y_{t+j}\}_{j=1}^{l}$ 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 $\{Y_{t+j}\}_{j=1}^{l}$ 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 $\alpha_{j+(w-1)}$).

B.3 The Number of $Z_{t+k_{ij}}$

Now we come back to (B.1) to find the Z terms in each of $\{Y_{t+j}\}_{j=1}^{l}$ 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 $\{Y_{t+j}\}_{j=1}^{l}$ has $q+1$ innovation terms $\{Z_{t+j}, Z_{t+j-1}, \dots, Z_{t+j-q}\}$. However, by repeated substitution, each $\{Y_{t+j}\}_{j=1}^{l}$ 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 $\{Y_{t+j}\}_{j=1}^{l}$ is equal to the number of innovation terms in the p previous observations, plus $q+1$.

B.4 The Coefficient of $Z_{t+k_{ij}}$

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 $k = 1, \dots, p$.

B.5 The Subscript of $Z_{t+k_{ij}}$

$t + k_{ij}$ denotes the subscript of Z for each i, j ($j = 1, \dots, l$ and $i = 1, \dots, n_j^{p+1}$). Each $\{Y_{t+j}\}_{j=1}^{l}$ 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 $\{Y_{t+j}\}_{j=1}^{l}$ are $\{j-q, j-1, \dots, j\}$. This is shown in Table 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 $\{Y_{t+j}\}_{j=1}^{l}$ simply are the subscripts of the innovation terms of p previous observations.

As a result, the aggregated process can be expressed as (18) with the associated parameters as defined in Table 1.

APPENDIX C- SIMULATION RESULTS FOR HORIZON-AGGREGATED FORECASTS FOR INAR(1), INMA(1) AND INARMA(1,1) MODELS

In this appendix, the forecast accuracy of the two forecasting methods (INARMA-Agg and INARMA-h) is compared in terms of MSE. The results are for the cases where data is produced by an INAR(1), an INMA(1) or an INARMA(1,1) process and the forecast horizon is $l = 6$ or $l = 9$.

Table C-1 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INAR(1) series when $l = 6$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.9534	0.9750	0.9961	0.9979	0.9534
$\alpha = 0.5, \lambda = 0.5$	0.8852	0.8976	0.9070	0.8783	0.8852
$\alpha = 0.1, \lambda = 1$	0.9515	0.9677	0.9929	0.9966	0.9515
$\alpha = 0.5, \lambda = 1$	0.8610	0.8545	0.8474	0.8302	0.8610
$\alpha = 0.1, \lambda = 3$	0.9444	0.9567	0.9811	0.9923	0.9444
$\alpha = 0.5, \lambda = 3$	0.7693	0.7365	0.6942	0.6742	0.7693
$\alpha = 0.1, \lambda = 5$	0.9211	0.9398	0.9677	0.9857	0.9211
$\alpha = 0.5, \lambda = 5$	0.9534	0.9750	0.9961	0.9979	0.9534

Table C-2 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INAR(1) series when $l = 9$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \lambda = 0.5$	0.6497	0.8698	0.9284	0.9868	0.9983
$\alpha = 0.5, \lambda = 0.5$	0.6973	0.8425	0.8887	0.9251	0.9189
$\alpha = 0.1, \lambda = 1$	0.6418	0.8723	0.9305	0.9829	0.9960
$\alpha = 0.5, \lambda = 1$	0.6924	0.8362	0.9046	0.9056	0.8920
$\alpha = 0.1, \lambda = 3$	0.6657	0.8574	0.8979	0.9717	0.9939
$\alpha = 0.5, \lambda = 3$	0.6235	0.7602	0.8016	0.8048	0.7732
$\alpha = 0.1, \lambda = 5$	0.6055	0.8669	0.9119	0.9677	0.9904
$\alpha = 0.5, \lambda = 5$	0.6384	0.7262	0.7532	0.7052	0.6736

Table C-3 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INMA(1) series when $l = 6$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.9996	0.9996	0.9999	1.0000	1.0000
$\beta = 0.5, \lambda = 0.5$	0.9982	0.9989	0.9992	0.9996	0.9999
$\beta = 0.9, \lambda = 0.5$	0.9969	0.9978	0.9984	0.9992	0.9998
$\beta = 0.1, \lambda = 1$	0.9997	0.9998	0.9999	1.0000	1.0000
$\beta = 0.5, \lambda = 1$	0.9986	0.9988	0.9992	0.9995	0.9999
$\beta = 0.9, \lambda = 1$	0.9966	0.9979	0.9983	0.9991	0.9998
$\beta = 0.1, \lambda = 3$	0.9999	0.9996	1.0000	1.0000	1.0000
$\beta = 0.5, \lambda = 3$	0.9974	0.9987	0.9992	0.9995	0.9999
$\beta = 0.9, \lambda = 3$	0.9948	0.9975	0.9981	0.9991	0.9998
$\beta = 0.1, \lambda = 5$	1.0002	0.9999	0.9998	1.0000	1.0000
$\beta = 0.5, \lambda = 5$	0.9976	0.9988	0.9992	0.9995	0.9999
$\beta = 0.9, \lambda = 5$	0.9945	0.9975	0.9980	0.9990	0.9998

Table C-4 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INMA(1) series when $l = 9$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\beta = 0.1, \lambda = 0.5$	0.9995	0.9998	1.0000	1.0000	1.0000
$\beta = 0.5, \lambda = 0.5$	0.9984	0.9992	0.9995	0.9997	0.9999
$\beta = 0.9, \lambda = 0.5$	0.9973	0.9986	0.9989	0.9995	0.9999
$\beta = 0.1, \lambda = 1$	0.9995	0.9999	0.9999	1.0000	1.0000
$\beta = 0.5, \lambda = 1$	0.9984	0.9993	0.9995	0.9997	0.9999
$\beta = 0.9, \lambda = 1$	0.9973	0.9985	0.9989	0.9994	0.9999
$\beta = 0.1, \lambda = 3$	0.9991	0.9998	0.9999	1.0000	1.0000
$\beta = 0.5, \lambda = 3$	0.9978	0.9994	0.9994	0.9997	0.9999
$\beta = 0.9, \lambda = 3$	0.9973	0.9985	0.9990	0.9994	0.9999
$\beta = 0.1, \lambda = 5$	1.0000	1.0000	0.9999	1.0000	1.0000
$\beta = 0.5, \lambda = 5$	0.9975	0.9987	0.9995	0.9997	0.9999
$\beta = 0.9, \lambda = 5$	0.9955	0.9976	0.9987	0.9994	0.9999

Table C-5 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INARMA(1,1) series when $l = 6$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	1.1280	1.1290	1.1129	1.0340	0.9975
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.9817	0.9902	0.9727	0.9676	0.9864
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.9291	0.8821	0.9278	0.9009	0.9044
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	1.0289	1.1009	1.0355	1.0200	0.9991
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.9652	1.0028	0.9657	0.9736	0.9857
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.8933	0.9272	0.8895	0.9066	0.9065
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.8930	1.0278	1.0238	1.0005	0.9966
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.9320	0.9472	0.9848	0.9668	0.9829
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.8281	0.8787	0.8837	0.8904	0.9029

Table C-6 $\text{MSE}_{Agg}/\text{MSE}_h$ of aggregated forecasts for INARMA(1,1) series when $l = 9$

Parameters	$n = 24$	$n = 36$	$n = 48$	$n = 96$	$n = 500$
$\alpha = 0.1, \beta = 0.1, \lambda = 0.5$	0.8270	1.1104	1.1042	1.0444	0.9984
$\alpha = 0.1, \beta = 0.9, \lambda = 0.5$	0.7848	0.9304	0.9364	0.9652	0.9898
$\alpha = 0.5, \beta = 0.5, \lambda = 0.5$	0.8773	0.8642	0.9091	0.9123	0.9389
$\alpha = 0.1, \beta = 0.1, \lambda = 1$	0.7756	0.9952	1.0110	1.0206	0.9981
$\alpha = 0.1, \beta = 0.9, \lambda = 1$	0.6699	0.9117	0.9455	0.9603	0.9901
$\alpha = 0.5, \beta = 0.5, \lambda = 1$	0.5531	0.8320	0.8976	0.9279	0.9372
$\alpha = 0.1, \beta = 0.1, \lambda = 5$	0.3814	0.8981	0.9941	1.0063	0.9967
$\alpha = 0.1, \beta = 0.9, \lambda = 5$	0.3441	0.8315	0.9185	0.9654	0.9862
$\alpha = 0.5, \beta = 0.5, \lambda = 5$	0.3577	0.8164	0.8840	0.9123	0.9332

References

1. Quenouille, M.H. Discrete autoregressive schemes with varying time intervals. Metrika 1958;1:21-27.
2. Amemiya, T. and R.Y. Wu. The effect of aggregation on prediction in the autoregressive model. Journal of the American Statistical Association 1972;67(339):628-632.
3. Brewer, K.R.W. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. Journal of Econometrics 1973;1(2):133-154.
4. Harvey, A.C. and R.G. Pierse. Estimating missing observations in economic time series. Journal of the American Statistical Association 1984;79(385):125-131.
5. Nijman, T.E. and F.C. Palm. Predictive accuracy gain from disaggregate sampling in ARIMA models. Journal of Business & Economic Statistics 1990;8(4):405-415.
6. Drost, F.C. and T.E. Nijman. Temporal aggregation of GARCH processes. Econometrica 1993;61(4):909-927.
7. Marcellino, M. Some consequences of temporal aggregation in empirical analysis. Journal of Business & Economic Statistics 1999;17(1):129-136.
8. Teles, P. and W.W.S. Wei. The use of aggregate time series in testing for Gaussianity. Journal of Time Series Analysis 2002;23(1):95-116.
9. Man, K.S. Linear prediction of temporal aggregates under model misspecification. International Journal of Forecasting 2004;20(4):659-670.
10. Tang, D. Managing finished-goods inventory under capacitated delayed differentiation. OMEGA: The International Journal of Management Science 2011;39(5):481-492.
11. Hosoda, T. and S.M. Disney. On variance amplification in a three-echelon supply chain with minimum mean square error forecasting. OMEGA: The International Journal of Management Science 2006;34(4):344-358.
12. Thury, G. and S.F. Witt. Forecasting industrial production using structural time series models. OMEGA: The International Journal of Management Science 1998;26(6):751-767.
13. Brännäs, K., J. Hellström, and J. Nordström. A new approach to modelling and forecasting monthly guest nights in hotels. International Journal of Forecasting 2002;18(1):19-30.
14. Freeland, R.K. and B.P.M. McCabe. Forecasting discrete valued low count time series. International Journal of Forecasting 2004;20:427-434.
15. Gourieroux, C. and J. Jasiak. Heterogeneous INAR(1) model with application to car insurance. Insurance: Mathematics and Economics 2004;34(2):177-192.
16. Brännäs, K. and S. Quoreshi. Integer-valued moving average modelling of the number of transactions in stocks. Applied Financial Economics 2010;20(18):1429-1440.
17. Sueutel, F.W. and K. van Harn. Discrete analogues of self-decomposability and stability. Annals of Probability 1979;7(5):893-899.
18. Mohammadipour, M. Intermittent demand forecasting with integer autoregressive moving average models 2009 Buckinghamshire New University, Brunel University.
19. Du, J.G. and Y. Li. The integer-valued autoregressive (INAR(p)) model. Journal of Time Series Analysis 1991;12(2):129-142.
20. McKenzie, E. Some ARMA models for dependent sequences of Poisson counts. Advances in Applied Probability 1988;20(4):822-835.
21. Neal, P. and T.S. Rao. MCMC for integer-valued ARMA processes. Journal of Time Series Analysis 2007;28(1):92-110.
22. Brännäs, K. and J. Hellström. Generalized integer-valued autoregression. Econometric Reviews 2001;20(4):425-443.

23. Jung, R.C. and A.R. Tremayne. Coherent forecasting in integer time series models. International Journal of Forecasting 2006;22:223-238.
24. Eaves, A.H.C. Forecasting for the ordering and stock-holding of consumable spare parts 2002 Lancaster University.
25. Al-Osh, M.A. and A.A. Alzaid. First order integer-valued autoregressive (INAR(1)) processes. Journal of Time Series Analysis 1987;8(3):261-275.
26. Bu, R., B. McCabe, and K. Hadri. Maximum likelihood estimation of higher-order integer-valued autoregressive processes. Journal of Time Series Analysis 2008;29(6):973-994.
27. Silva, M.E. and V.L. Oliveira. Difference equations for the higher-order moments and cumulants of the INAR(1) model. Journal of Time Series Analysis 2004;25(3):317-333.
28. Freeland, R.K. and B.P.M. McCabe. Asymptotic properties of CLS estimators in the Poisson AR(1) model. Statistics and Probability Letters 2005;73(2):147-153.
29. Cox, D.R. Prediction by exponentially weighted moving averages and related methods. Journal of the Royal Statistical Society. Series B 1961;23:414-422.
30. Tiao, G.C. and D. Xu. Robustness of maximum likelihood estimates for multi-step predictors: the exponential smoothing case. Biometrika 1993;80:623-641.
31. Kang, I.-B. Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data. International Journal of Forecasting 2003;19:387-400.
32. Syntetos, A.A. and J.E. Boylan. The accuracy of intermittent demand estimates. International Journal of Forecasting 2005;21:303-314.
33. Latour, A. Existence and stochastic structure of a non-negative integer-valued autoregressive process. Journal of Time Series Analysis 1998;19(4):439-455.
34. Teunter, R.H. and L. Duncan. Forecasting intermittent demand: a comparative study. Journal of the Operational Research Society 2009;60(3):321-329.
35. da Silva, C.G., et al. An interactive decision support system for an aggregate production planning model based on multiple criteria mixed integer linear programming. OMEGA: The International Journal of Management Science 2006;34(2):167-177.