CLINT PENNINGS

Advancements in Demand Forecasting

Methods and Behavior

Advancements in Demand Forecasting: Methods and Behavior

Advancements in Demand Forecasting: Methods and Behavior

Verbeteringen in het voorspellen van vraag: methoden en gedrag

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Thursday 10 November 2016 at 13:30 hours

by

Clint Leonardus Petrus Pennings
born in 's-Hertogenbosch, the Netherlands.

Erasmus University Rotterdam

Doctoral Committee

Promotors:	Prof.dr. L.G. Kroon
	Prof.dr.ir. H.W.G.M. van Heck
Other members:	Prof.dr. A.G. de Kok
	Prof.dr.ir. M.B.M. de Koster
	Prof.dr. A.A. Syntetos
Copromotor:	Dr. J. van Dalen

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://repub.eur.nl/pub/

ERIM PhD Series in Research in Management, 400
ERIM reference number: EPS-2016-400-LIS
ISBN 978-90-5892-465-0

This research was funded by the Dutch Institute for Advanced Logistics (DINALOG) as part of the 4C4More project.

Design: PanArt, www.panart.nl

This publication (cover and interior) is printed by Tuijtel on recycled paper, BalanceSilk®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Ecolabel, FSC®, ISO14001.
More info: www.tuijtel.com

Acknowledgments

Ever since I first entered university, I have met many people who inspired me to study, challenge, and teach, and who proved to be excellent role models. I will never forget their lessons. But most importantly, I will never forget them. Academia is at its finest when you are able to experience it as a single collaborative effort. Being part of such a community is both very enjoyable and humbling.

My deepest gratitude goes to my daily supervisor, Jan van Dalen, whom I hold in the highest regard. Students—whether PhD, Master, or Bachelor—always receive his utmost attention and support, and he never ceases to challenge and question. I benefited immensely from his guidance, and greatly enjoyed working with him in research and in teaching. His magnanimity is unparalleled.

I am also grateful to my promotors, Leo Kroon and Eric van Heck, for their invaluable help and time. I greatly miss Leo, and I regret wanting to wait for the defense to tell him how grateful I am. Eric is great in framing studies in a wider context, which both motivates you to do better, and to do more.

The research conducted in this thesis is part of the research project 4C4More, sponsored by DINALOG. Without Ton de Kok, the head of this project, I would not have had the opportunity to perform this research. More importantly, without him the project research meetings would not have been half as enjoyable. Thanks also to the other members of the project, Jos, Kasper, Matthew, Robbert, Simon, and Sjoerd for sharing ideas and feedback.

I would also like to thank the other members of the inner thesis committee, René de Koster and Aris Syntetos, for assessing the thesis and for their comments and thoughts. René, I very much appreciate all of your help. Aris, I am grateful for the critical reflections during conferences and workshops, and that you were able to travel to Rotterdam.

I also thank my co-authors, Erwin van der Laan, Laurens Rook, and Stefanie Protzner, for the fruitful collaboration; my teachers prior to starting my PhD, especially Paul Franssen; my colleagues, especially the ones with whom I taught courses; Carmen, Cheryl, and Ingrid for their unwavering assistance; the ERIM office, especially Kim, Miho, and Tineke; and many friends: Irina, Joris, Judith, Konstantina, Panos, Sarita, Xiao, and Yixin. Thomas, thanks for the great time. Luuk, Nick, Paul, and Wouter: it has been an absolute pleasure, thank you for being the ones.

Thanks to my parents and family for their support, and especially to my wife Willemijn, for her love, but mostly for being awesome.

Rotterdam, October 2016
Clint Pennings

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Forecasting is necessary to support various activities in supply chains—inventory control, budgeting, production and distribution planning—as a result of demand uncertainty (Danese and Kalchschmidt, 2011). Retailers use forecasts as input for sales, inventory and order decisions, suppliers for production and procurement decisions, and distributors for capacity allocation decisions. In practice, forecast errors are relatively large (Hughes, 2001), which negatively affects operational performance (Danese and Kalchschmidt, 2011; Enns, 2002; Ritzman and King, 1993; Zhao and Xie, 2002). Reducing or minimizing these forecast errors, following a profound understanding of their origins, is therefore of central importance.

Supply chain actors traditionally produce forecasts on their own, using data generated by their transactions and operational decisions and available to them through their own databases. But research has found that substantial savings can be attained, mainly in terms of reduced inventory, from sharing information and forecasts between actors in the supply chain (Huang et al., 2003). Aviv (2001) concludes that sharing forecasts increases, and never decreases, forecast accuracy, as errors in forecasts propagate upstream through orders, distorting the basis on which decisions are made. Moreover, sharing information reduces the bullwhip effect (Chen et al., 2000), dampening the increase in demand variability observed by actors upwards in the chain (Lee et al., 1997).

Typically, analyses depend on stylized demand processes, in which demand d_t at time period t follows an autoregressive process (AR) for various values of parameter ϕ:

$$d_t = \phi d_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{1.1}$$

The form of the underlying process is assumed to be stable and known, including the values of the parameters, to the retailer only or all parties in the supply chain (Aviv, 2002; Chen and Lee, 2009; Gaur et al., 2005; Ha and Tong, 2008; Lee et al., 2000).

Raghunathan (2001) shows that sharing forecasts is redundant if the process is known, because this allows the manufacturer to infer the demand pattern from the order history alone. This insight led to research further exploring under what conditions the demand process can be inferred from order patterns. The demand process in Equation (1.1) is generalized to the following ARMA process, which consists of an autoregressive (AR) and moving average (MA) component:

$$d_t = c + \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{1.2}$$

Zhang (2004) and Gilbert (2005) show that this type of demand process generates an ARMA order history, when order decisions are made based on a rational order-up-to policy. If forecasts with minimum mean squared error are used by actors at each stage of the supply chain, and the orders are transmitted immediately, there is no need to share the demand information between supply chain actors (Gilbert, 2005). Demand information is always less valuable if actors in the supply chain have rational ordering policies: for example, just when the demand information is most valuable to the manufacturer, the retailer is likely to transmit this information by submitting an order (Cachon and Fisher, 2000). Even when there is a small delay in ordering, the manufacturer can, in many cases, still infer demand from orders (Gaur et al., 2005).

These results critically depend on the assumption of a rational order-up-to policy. Also, these results critically depend on the forecasting process. Ali et al. (2012) show that if single exponential smoothing, a popular forecasting method in practice, is used by a supply chain actor given the demand described in Lee et al. (2000), demand cannot be inferred by actors further up the chain. Given more realistic assumptions concerning ordering policies and forecasting processes, Ali and Boylan (2011) conclude that it is not possible to infer demand.

The discussion of ordering policies and forecasting processes has evolved within stylized data-generating processes. Research widely concludes that the benefit of information and forecast sharing is highly dependent upon and sensitive to the specification of the demand process (Ali et al., 2012; Babai et al., 2013; Bourland et al., 1996; Gaur et al., 2005; Lee et al., 2000; Zhao et al., 2002). The formulation of the data-generating process under study is critical for the results to hold.

However, specifying a stylized demand process for practice is questionable, as it is doubtful whether demand processes are ever stable in practice. Promotions are an important case in which stylized demand processes are not adequate approximations due to the substantial uncertainty of demand. During promotions, the larger forecast errors and large increases in volume translate into higher inventory costs, undermining the desired profit which is targeted by the promotion. Promotions are the main cause of out-of-stocks, excess inventory, and unplanned logistics costs (Wiehenbrauk, 2010). Iyer and Ye (2000) show that the profitability of a promotion decreases in conjunction with the forecast accuracy of demand during promotions—even to such an extent that it may be more beneficial to avoid promotions all together. In addition, demand is often distorted by outside influences, like activities and promotions of competitors, which can have a severe impact (Wiehenbrauk, 2010). Volumes peak during one's own promotions, but drop during competitors' promotions, and forecast accuracy suffers drastically in both of these periods.

Not surprisingly, the stylized models for the demand process have been criticized as 'implausible,' and for 'lack[ing] any empirical foundation' (Fildes et al., 2008, p. 1162). In practice, companies are skeptical about how well demand can be forecast, and question the forecasting process itself: they doubt whether a level of accuracy can be attained to justify the effort involved (Hughes, 2001). This difference between research and practice is not surprising, as the assumption of a stable demand process in the models does not take the uncertainty of forecasting modeling into account (Chatfield, 1996). Even if there is a stable demand process, companies are not necessarily able to identify this process based on historical data.

In the case of Aviv (2001, 2007), forecast information and forecasting capability are conflated: more information immediately translates into a higher forecast accuracy. Moreover, the forecasting processes are treated as identical. However, the forecasting horizons and units are different for the forecasting tasks of retailers and manufacturers. Retailers often have to forecast consumer demand for short-term stocking decisions, whereas manufacturers have to forecast the actual orders placed by the retailer for long-term production decisions.

The distinction between information and capability is important in view of the limitations in capability, in terms of forecast model formulation and estimation, of retailers and manufacturers (Småros, 2007). Retailers and manufacturers are generally unable to adequately handle all the information currently at their disposal, not able to separate demand signals from noise, and their limited forecasting capabilities impede them from using more information to improve forecast accuracy.

Generally, companies lack the knowledge, expertise and training in the field of forecasting to validly support decision-making (Hughes, 2001). The situation has even become worse, as the level of knowledge and forecast accuracy have decreased over time (McCarthy et al., 2006). Davis and Mentzer (2007) observe a gap between theory and practice in terms of forecasting capability, and consider this a significant issue.

Hence, researchers call for analyses that extend beyond the technical side of forecast generation and focus on how the forecasting process is managed and organized (Armstrong, 1987; Danese and Kalchschmidt, 2011). There is a lack of performance evaluation and management of forecasting processes at companies, and a blurred distinction between forecasts, plans, and goals (Moon et al., 2003). Moreover, forecasts generated by forecasting methods are not directly used. The use of judgment for generating and adjusting forecasts is often preferred and widely used (Hughes, 2001; Lawrence et al., 2000).

1.1 Motivation

The problems associated with forecasting capability and forecasting process management motivate the studies presented here. Instead of an analytical approach based on stylized models, the studies in this thesis analyze and draw conclusions based on empirical data collected from industry. Rather than focusing on a whole supply chain, the scope is restricted to manufacturers, as improvements in the forecasting capability of manufacturers not only benefit their own forecasts, but are most beneficial to the chain as a whole (Aviv, 2001, 2007). In the order they are presented, the studies gradually extend the use of information and move toward the actual use of forecasts, incorporating managerial behavior. The issues encountered are generalizable to a wider supply chain setting.

Forecasting capability here refers to the entire forecasting process at a company, which can broadly be divided into two stages: (1) using various forecasting models and methods to generate forecasts, and (2) using judgment to generate forecasts or adjust previously generated forecasts by various parties in the company (Fildes et al., 2008). Both of these stages are explored in this thesis.

The first stage, relating to models and methods, is explored in two different studies, which can be summarized as exploiting already available information, without context-specific assumptions, to achieve substantial gains. The first study exploits the available information about demand intermittence by generalizing existing forecasting methods to better model demand. Existing methods either ignore a dependency between the time between orders and order size, or focus on the risk of inventory obsolescence. This limited scope is costly in terms of inventory and financial performance. The second study also generalizes existing forecasting methods and supersedes the traditional discussion of top-down versus bottom-up methods, by examining how the hierarchy of products is used in forecasting. Stock-keeping units (SKUs) naturally group together in a hierarchy going from the bottom, with individual sales per product, through several intermediary levels, denoting sales for groups of related products at increasingly general aggregation levels, such as product groups and categories, to the top of the hierarchy, which lists total sales. Two commonly used approaches in practice and research start from opposite ends of the hierarchy to generate forecasts for all series: bottom-up forecasting and top-down forecasting. Research stretches over three decades with mixed results as to preference for either bottom-up or top-down forecasting approaches. Both of these approaches imply a loss of information because the scope is restricted to separate and independent initial forecasts. However, by generating joint forecasts for a group of products directly, information is better used which translates to superior performance. This approach explicitly incorporates product dependencies, such as complementarity of products and product substitution, which are otherwise ignored. Whereas the first study exploits available information for each SKU separately, the second study does so by considering SKUs in groups and hierarchies, expanding the scope of the forecasting models.

The second stage, relating to the use of judgment, is explored in the final two studies, which focus on forecasting processes at companies, which extend beyond applying forecasting methods and models. The studies use behavioral experiments to provide insights into how forecaster behavior systematically differs and how this and the design of the forecasting process affect performance. Fildes et al. (2008) observe that judgmental forecasting is central to the forecasting processes at many companies, and directly affects supply chain performance (Syntetos et al., 2011, 2010). Judgmental forecasting is often used to capitalize on valuable tacit or domain-specific knowledge which is not captured by models (Fildes et al., 2008). Yet, judgmental forecasting introduces many inherent biases of human decision-making (Lawrence et al., 2006), which can adversely affect the forecast even more than the use of tacit information improves it (Lawrence et al., 2006). Behavioral operations research examines how to manage judgment, an “indispensable component” of forecasting (Lawrence et al., 2006, p. 493). However, this is generally done at an aggregate level. We examine differences in forecasting behavior by modeling individual forecasting behavior, instead of a demand process. We demonstrate that forecasting behavior differs systematically between individuals to the extent that we discern two markedly different types of forecasters, chasers and smoothers. We also examine the influence of roles and incentives, and trace the extent to which forecasters intentionally adjust their forecasts. These insights have ramifications for hiring forecasters and orchestrating forecasting processes.

An aspect addressed in all of the studies in this thesis is that the focus is not restricted to improving forecast accuracy. There are many different ways of evaluating forecast accuracy (e.g. Hyndman and Koehler, 2006). However, higher forecast accuracy does not always translate into operational gains (Fildes et al., 2008). Recent papers have evaluated forecasting methods in terms of their resulting improvement for inventory management (e.g. Syntetos and Boylan, 2006; Syntetos et al., 2010). This insight has been adopted here in addition to directly evaluating forecast accuracy.

1.2 Contributions

In this section the content of the research described in Chapters 2 to 5 is summarized.

Exploiting Elapsed Time for Managing Intermittent Demand for Spare Parts
Chapter 2 presents an intermittent demand forecasting method that conditions on the elapsed time since the last demand occurrence to anticipate incoming demand and shows, using empirical data, that this can substantially reduce both stock investment and lost revenue for spare parts. We extensively benchmark our method against existing forecasting and bootstrapping methods on forecast accuracy and inventory performance and demonstrate that its performance is robust under general conditions. Existing forecasting methods either do not change the forecast after a period of zero demand, ignoring all forms of cross-correlations, or adjust the forecast downwards, addressing only the specific case of inventory obsolescence and not the general forms of cross-correlations observed in empirical data. All methods ignore the fact that activities at the source of the demand, such as aggregation of demand, preventive and corrective maintenance, can lead to a positive relation between demand size and inter-arrival time of demand occurrences. By anticipating incoming demand and not exclusively focusing on spare parts obsolescence, our method offers substantial financial gains.

This chapter demonstrates that, even without context-specific knowledge and assumptions, and even if there is very little information available, currently available information can still be used to substantially improve performance. By extending forecasting methods from literature, the forecasting capability of manufacturers is directly increased. This chapter treats products independently, as the little available information is not enough to estimate dependencies between products.

Integrated Hierarchical Forecasting
Chapter 3 looks into generating forecasts for product groups, and specifically examines product dependencies ignored in practice. Forecasts are often made at various levels of aggregation of individual products, which combine into groups at higher hi-

erarchical levels. We provide an alternative to the traditional discussion of bottom-up versus top-down forecasting by examining how the hierarchy of products can be exploited when forecasts are generated. Instead of selecting series from parts of the hierarchy for forecasting, we explore using all the series. Moreover, instead of using the hierarchy after initial forecasts are generated, we consider the hierarchical series as a whole to instantaneously generate forecasts for all levels of the hierarchy. Our integrated approach explicitly incorporates product dependencies, such as complementarity of products and product substitution, which are otherwise ignored. A simulation study, comparing and contrasting existing approaches from literature under possible cross-correlations and dependencies, shows the conditions under which an integrated approach is advantageous. An empirical study shows the substantial gain, in terms of forecast performance as well as inventory performance, of generalizing the bottom-up and top-down forecasting approaches to an integrated approach. Specifically, the gains for inventory performance can be as much as a 39% reduction in stock investment. The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of forecasting for manufacturers.

This chapter further extends forecasting methods from literature, by superseding the discussion about top-down versus bottom-up forecasting approaches, by proposing an integrated forecasting approach. This approach translates into a substantial financial gain and extends the forecasting capability of manufacturers. It simplifies and consolidates the forecasting process by generating forecasts for multiple SKUs at once, instead of generating separate forecasts for each SKU. The chapter shows how a forecasting method can perform better, purely based on historic information, but does not include managerial and forecaster behavior, and does not consider how forecasts are generated in collaboration between multiple departments.

Chasers, Smoothers and Departmental Biases: Heterogeneity in Judgmental Forecasting

Chapter 4 demonstrates that forecasting behavior differs systematically between individuals to the extent that we discern two markedly different types of forecasters. One is characterized by overreaction to forecast errors and might be labeled chasers, while the other is characterized by underreaction to forecast errors, and might be labeled smoothers. Extending the models used in earlier behavioral experiments, our approach relies on wavelets and state space modeling to incorporate forecasting heterogeneity. We demonstrate that contextual biases can only be meaningfully explored after controlling for the forecaster's inclination towards chasing or smoothing. We further show that departmental biases persistently impact judgmental forecasting, even if forecasts are constructed to be free of intentional biases. Our findings have important repercussions for theory building based on evidence derived from aggregate results, but also have practical relevance for training and hiring of forecasters, and orchestrating forecasting processes in companies.

By shifting the attention to modeling forecaster behavior, instead of the demand process, this chapter gives insight into how forecasters behave, how they are influenced by context, and how this impacts performance. Forecasting behavior and its ramifications are to a large extent unintentional. In addition to these sources of unintentional errors, different incentives and departments in a company possibly affect intentional errors.

Coordinating Judgmental Forecasting: Coping with Intentional Biases
Chapter 5 examines intentional biases, an overlooked research area, that arise due to the influence of different departmental roles and incentives in the forecasting process. Through an experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, we examine the effects of roles, incentives, and various weighing schemes on behavior and performance. We find that roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We test the claim that an accuracy-weighted scheme can remove unintentional biases, and conclude that though this halves these biases, it does not fully remove them. Finally, we observe that a weighing scheme that explicitly corrects biased inputs shows great promise in reducing intentional as well as unintentional biases. In our experiment, this scheme reduces biases by 35%. This study shows the importance of disentangling the two sources of biases for research, and our insights have substantial ramifications for the design of the forecasting process in terms of coordination mechanisms and incentives by quantifying the impact of roles and incentives.

This chapter shows the impact of forecasting process design on performance, and how people intentionally adapt their behavior under commonly used schemes, which affects performance.

The research in all of these chapters leads to an increase in forecasting capability by extending the forecasting methods and models available to companies, and by showing how the forecasting process, in which these methods and models are embedded, can be improved.

1.3 Authorship

The majority of the work in this thesis has been done independently by the author. The author was responsible for formulating the research questions, studying relevant literature, conducting the analyses, formulating and implementing the models, analyzing the results, and writing the chapters. For each chapter, discussions with co-authors and promoters, especially with Jan van Dalen, led to substantial improvements. The data of Chapter 2 was made available by professor Aris Syntetos and Erwin van der Laan. The methodology section in this chapter is based on prior work by Jan van Dalen and Erwin van der Laan. The data of Chapter 3 was made available by the company. The setup of the behavioral experiments in Chapters 4 and 5 were outlined in collaboration with Jan van Dalen, Laurens Rook and Stefanie Protzner, and programmed by the author. Also, the author conducted the data collection sessions with practitioners.

1.4 Outline

The remainder of this thesis consists of the chapters described in Section 1.2. Chapters 2 and 3 present the work on improving the forecasting models and methods by modeling the data-generating process for demand and exploiting available, but currently ignored, information. Chapter 4 and 5 present the work on modeling individual forecaster behavior and the effect of forecasting process design on performance. Finally, Chapter 6 summarizes results, presents conclusions, and contains recommendations for future research.

Chapter 2

Exploiting Elapsed Time for Managing Intermittent Demand for Spare Parts

Co-authors: J. van Dalen and E. van der Laan

Abstract

We present an intermittent demand forecasting method that conditions on the elapsed time since the last demand occurrence to anticipate incoming demand and show, using empirical data, that this can substantially reduce both stock investment and lost revenue for spare parts management. We extensively benchmark our method against existing forecasting and bootstrapping methods on forecast accuracy and inventory performance and demonstrate that its performance is robust under general conditions. Our method is the first to incorporate that activities at the demand side, such as aggregation of demand, preventive and corrective maintenance, can lead to a positive relation between demand size and inter-arrival time of demand occurrences. By anticipating incoming demand, our method offers substantial financial gains.

Keywords: forecasting, spare parts, intermittent demand, Croston, bootstrap.

2.1 Introduction

Spare parts management is of great business value and is important for competitive success (Cohen and Lee, 1990). It is often applied to large numbers of stock keeping units (SKUs), ranging into the thousands. Service levels have to be met, while stock investments, which can represent a large capital, are curtailed. Forecasts drive inventory decisions, directly affecting stock investments and customer satisfaction, but are challenging to generate in the case of spare parts. Johnston et al. (2003) describe a company with 50,000 different SKUs of which the products with intermittent demand represent 87% of the total stock value and 60% of the value of sales. Glueck et al. (2011) survey the service and spare parts management activities of manufacturing companies and report that forecast accuracies are poor, and that almost "70% of the manufacturers surveyed are unable to report on the forecast accuracy for their service and parts activities" (p.28).

Spare parts are especially challenging to forecast because demand is typically intermittent with substantial and variable periods of time between demand occurrences. As a consequence, variability can occur not only in the demand size, but also in the inter-arrival time of demand occurrences. In spare parts management, inter-arrival times of more than a year are no exception (e.g. see data descriptive statistics in section 4.2). Also, when demand arrives, it can be for large quantities of items, even ranging into the thousands. This particular setting requires special forecasting methods, as demonstrated by Croston (1972), whose forecasting method is still widely used today.

Croston proposed an intermittent demand forecasting method that distinguishes between the demand size and the inter-arrival time of subsequent demand occurrences. The method assumes independence between the demand size and the inter-arrival time of demand. However, empirical data can exhibit substantial cross-correlations between these two (Willemain et al., 1994). Simulations show that ignoring these cross-correlations adversely affects the service level (Altay et al., 2012). Since Boylan and Syntetos (2007)'s claim that "no methods have yet been published to address the general case of non-independence" (p.513), newer methods have been developed to relax the initial assumption, starting with Teunter et al. (2011). These recent methods specifically address inventory obsolescence, which the other methods ignore. Yet no method addresses more general forms of cross-correlations observed in empirical data. More importantly, behavior on the demand side, which characterizes the demand process, is ignored. Inderfurth and Kleber (2013) provide an example of such behavior, where a final large quantity of parts is ordered at the end of the life cycle. Wang and Syntetos (2011) are the first to characterize the maintenance driven models, such as preventive maintenance and corrective maintenance, as a source of generating intermittent demand. Later work, such as by Romeijnders et al. (2012), assumes that maintenance schemes drive demand and are the source of variability in the inter-arrival times and the demand size. If demand is indeed generated by this behavior, the independence of the size and the inter-arrival time is clearly violated, but not in a way captured by methods that incorporate obsolescence: any period without a demand occurrence should then lead to a higher, rather than lower, expectation of demand.

If available, extra information can be used to improve inventory management of spare parts (Li and Ryan, 2011). However, when customers are external parties, context-specific information, such as ordering policies and maintenance schemes of buyers, is often unavailable. In most cases, the only available information is a short history of previous demand from which little can be inferred due the intermittent nature of the demand.

Our main contribution is that we develop a method that incorporates the overlooked case of positive cross-correlation between inter-arrival times and demand sizes to anticipate incoming demand. Using empirical data, we show that our method substantially reduces both stock investment and lost revenue for spare parts management. We extensively benchmark our method against existing forecasting and bootstrapping methods on forecast accuracy and inventory performance, and show that its performance is robust under general conditions. Our insights contribute to the spare parts and inventory management literature in general, and specifically to the literature concerned with improving forecast accuracy and inventory performance for spare parts with an intermittent demand pattern (e.g. Altay et al., 2012; Boylan et al., 2008; Boylan and Syntetos, 2007; Croston, 1972; Snyder et al., 2012; Syntetos and Boylan, 2001, 2005, 2006; Syntetos et al., 2012; Teunter et al., 2011; Willemain et al., 2004).

The remainder of this chapter is organized as follows. In Section 2.2 we provide an overview of the relevant literature about intermittent demand forecasting methods and their underlying assumptions, specifically with respect to time dependence. In Section 2.3, we propose a general model and formulate a specific application to address the time dependence. In Section 2.4 we describe our empirical data and the basis on which we compare the various methods in terms of forecast accuracy and inventory performance. Section 2.5 lists the results and their implications, while Section 2.6 concludes and gives suggestions for future research.

2.2 Theoretical background

This section reviews the available methods for forecasting the intermittent demand of spare parts. Methods are often classified as either parametric or non-parametric. The group of parametric approaches mostly consists of adjustments to Croston's method. In the non-parametric group, various forms of the bootstrap are most prominent.

Emphasis is given to the time dependence in the forecasting methods and the current state of research.

2.2.1 Forecasting methods

The demand for most spare parts exhibits large variation in the inter-arrival times between the demand occurrences. Also, demand volumes are often seen to vary considerably. The popular forecasting method single exponential smoothing (SES) only captures the variation in the demand size as:

$$\hat{d}_{t+1|t} = \hat{d}_{t|t-1} + \alpha(d_t - \hat{d}_{t|t-1})$$

where $\hat{d}_{t+1|t}$ denotes the forecast of demand for period $t+1$ made at time t, d_t denotes the observed demand at time t, and α is a smoothing parameter constrained as $0 \leq \alpha \leq 1$. Croston (1972) shows that if SES is used to forecast intermittent demand, the forecast is lowest just before a demand occurrence, and highest just after it. As an alternative approach, Croston proposes to smooth the demand size s_t and the inter-arrival time i_t separately, where i_t denotes the number of periods since the last demand occurrence. This method is widely used today and is becoming more popular (Boylan and Syntetos, 2007). At the end of time period t, if no demand has occurred ($s_t = 0$) the forecast made at the end of time $t-1$ remains unchanged ($\hat{d}_{t+1|t} = \hat{d}_{t|t-1}$), but if demand does occur ($s_t > 0$) then the forecasts for $t+1$ are updated:

$$\hat{s}_{t+1|t} = \begin{cases} \hat{s}_{t|t-1} & \text{if } s_t = 0 \\ \hat{s}_{t|t-1} + \alpha_0(s_t - \hat{s}_{t|t-1}) & \text{if } s_t > 0 \end{cases}$$

$$\hat{i}_{t+1|t} = \begin{cases} \hat{i}_{t|t-1} & \text{if } s_t = 0 \\ \hat{i}_{t|t-1} + \alpha_1(i_t - \hat{i}_{t|t-1}) & \text{if } s_t > 0 \end{cases}$$

for given smoothing parameters α_0 and α_1. The use of separate smoothing parameters is a later suggestion by Schultz (1987). The demand forecast results from the combination of the two separate forecasts:

$$\hat{d}_{t+1|t} = \frac{\hat{s}_{t+1|t}}{\hat{i}_{t+1|t}}$$

As the demand size is assumed to be independent of the elapsed time, the demand forecast remains the same in periods between demand occurrences.

Croston's method is biased, as $E(d_t) = E(s_t/i_t) \neq E(s_t)/E(i_t)$ (Syntetos and Boylan, 2001). Several modifications of Croston's method have been proposed to address this (Levn and Segerstedt, 2004; Shale et al., 2006; Snyder, 2002; Syntetos and Boylan, 2001, 2005, 2006). Though some of the variants perform better than the original (Syntetos and Boylan, 2006), others overcompensate and have an even stronger bias by forecasting too low (Teunter and Sani, 2009). The adjustment proposed by Syntetos and Boylan (2005), hereafter referred to as SBA, has the most empirical support and incorporates a correction factor to reduce the forecast:

$$\hat{d}_{t+1|t} = \left(1 - \frac{\alpha_1}{2}\right) \frac{\hat{s}_{t+1|t}}{\hat{i}_{t+1|t}}$$

An extension to Croston's method has been proposed based on the risk of inventory obsolescence, as the forecast of Croston's method does not change if there are no more demand occurrences. Teunter et al. (2011) (TSB) propose to update the probability that demand occurs $\hat{p}$ instead of the inter-arrival time in every period:

$$\hat{s}_{t+1|t} = \begin{cases} \hat{s}_{t|t-1} & \text{if } s_t = 0 \\ \hat{s}_{t|t-1} + \alpha_0(s_t - \hat{s}_{t|t-1}) & \text{if } s_t > 0 \end{cases}$$

$$\hat{p}_{t+1|t} = \begin{cases} (1-\alpha_1)\hat{p}_{t|t-1} & \text{if } s_t = 0 \\ (1-\alpha_1)\hat{p}_{t|t-1} + \alpha_1 & \text{if } s_t > 0 \end{cases}$$

$$\hat{d}_{t+1|t} = \hat{p}_{t+1|t}\hat{s}_{t+1|t}$$

Smoothing constant α_1 reduces the demand probability, and so also the demand forecast, in every period in which there is no demand, unless it is strictly equal to 0. The demand forecast $\hat{d}_{t+1|t}$ is then dependent on the elapsed time since the last demand occurrence at every period t. The smoothing using constant α_1 makes the method more similar to SES, because the forecast is again lowest just before and highest just after a demand occurrence.

Snyder et al. (2012) use a selection of count distributions (Poisson, negative binomial, and a hurdle shifted Poisson) to forecast intermittent demand, and add two dynamic specifications so that the mean of the demand distribution can change over time. The first corresponds to a stationary autoregressive model:

$$\hat{d}_{t+1|t} = (1-\phi-\alpha)\mu + \phi\hat{d}_{t|t-1} + \alpha d_t$$
$$\mu > 0,\ \phi > 0,\ \alpha > 0,\ \phi + \alpha < 1$$

where μ is the stationary mean, and α and ϕ are parameters. The constraints $\alpha > 0$ and $\phi + \alpha < 1$ imply that the updated forecast is a convex combination of the stationary mean, the previous forecast and the observed demand in each period. The second corresponds to an integrated moving average, or local level model:

$$\hat{d}_{t+1|t} = \delta \hat{d}_{t|t-1} + \alpha d_t$$
$$\delta > 0,\ \alpha > 0,\ \delta + \alpha = 1$$

where the constraints $\alpha > 0$ and $\delta + \alpha = 1$ imply that the updated forecast is a convex combination of the previous forecast and the observed demand. In these dynamic models, there is a dependence between the elapsed time and the demand size in the model formulation. They mention that "demand for spare parts may increase over time as the machines age and then decline as they fail completely or are withdrawn from service" (Snyder et al., 2012, p.486). An increase or decrease in the demand level is similarly captured in Croston's method, which also smoothes the demand size. However, similar to the approach of Teunter et al. (2011), the dynamic specification is able to incorporate inventory obsolescence when demand does not occur, which Croston's method is unable to do. For both the stationary autoregressive model and the local level model, the updated forecast becomes lower when no demand occurs. The approach of Snyder et al. (2012) is conceptually similar to SES for intermittent demand: its demand forecast is also lowest just before and highest just after demand occurs.

Prominent alternatives to the parametric methods are the methods based on bootstrapping, in which the empirical distribution of demand for SKUs is used directly to approximate the demand distribution. Willemain et al. (2004) use a two state, first-order Markov process to incorporate autocorrelation. The first step in this approach is to estimate the state transition probabilities from the historical demand series. Forecasts of demand occurrences during the lead time are conditional on whether demand just occurred ($s_t > 0$) or not ($s_t = 0$) at the moment of forecasting ($\hat{d}_{t+1|t}$). If demand no longer occurs, as in the case of obsolescence, the transition probability of going from the state with zero demand to a state of positive demand becomes ever smaller, so the demand forecast decreases. The Willemain et al. (2004) bootstrap, though nonparametric, is similar to the discussed alternatives to Croston's method in its application, because the demand forecast decreases in periods when no demand occurs.

Many authors conclude that more empirical studies are needed to evaluate the performance of the bootstrap and the various parametric methods, as few studies compare all of these based on measuring inventory performance (Gardner and Koehler, 2005; Syntetos et al., 2012; Teunter et al., 2011).

2.2.2 Elapsed time dependence

In Croston's method, the inter-arrival time and the demand size are independent by assumption. However, empirical data can exhibit substantial cross-correlations between these two (Willemain et al., 1994). This is a concern, because simulations show that the service level is affected when these cross-correlations are not taken into account, although this has not yet been tested on empirical data (Altay et al., 2012). Boylan and Syntetos (2007) state that "no methods have yet been published to address the general case of non-independence" (p.513). Since their statement, many methods have been developed in which a possible dependence is taken into account in the formulation of the expected demand for the specific case of inventory obsolescence. A longer inter-arrival time leads to a smaller expected demand in these methods. The general case of cross-correlation between the demand size and the inter-arrival time has, however, not been addressed.

Porras and Dekker (2008) note that their bootstrap method, which constructs a histogram of demands over the lead time without sampling, can capture the fixed demand intervals arising from preventive maintenance, which is important because the source of intermittent demand for spare parts is rarely explored. Wang and Syntetos (2011) were the first to characterize the maintenance driven models, such as preventive maintenance and corrective maintenance, as the source of the intermittent demand generation process in simulations. Romeijnders et al. (2012) consider the maintenance scheme as the source of variability in the inter-arrival time and the demand size and propose a method that takes the type of component for which a spare part is needed into account. In the setting of maintenance schemes, independence of demand size and inter-arrival time is clearly violated. The methods discussed so far cannot capture the relation exhibited in a maintenance setting—we have seen that there is no conceptual difference in the methods of Snyder et al. (2012) and Teunter et al. (2011).

A classification of SKUs, based on particular demand patterns, is available to select the best applicable forecasting method (e.g. Boylan et al., 2008). According to Syntetos and Boylan (2005, p.12), "the two key characteristics that have been shown to be collectively sufficient for defining intermittent demands are the inter-demand interval and the squared coefficient of variation of the demand sizes"; see also Syntetos et al. (2005). However, they also note that "key issues remaining in this area relate to [...] the further development of robust operational definitions of intermittent demand for forecasting and stock-control purposes" (Syntetos et al., 2012, p.3). Based on the discussion so far, cross-correlation seems a likely candidate to further capture essential aspects of the variation in demand between parts.

2.3 Proposed model

This section formalizes the intermittent demand process, and provides a general formulation of the intermittent demand model. This general model is then applied to the specific case in which the expected demand is proportional to a geometrically-distributed inter-arrival time.

2.3.1 General formulation

The objective is to estimate the expected total demand $D_{L,t}$ at time period t for some product or part during the next L time periods. The (expected) demand during the forecast period is by definition equal to the average of the expected demand for a given number of demand occurrences m, $D_{L,t|M=m}$, weighted with the probability of m demand occurrences $P(M = m)$, and aggregated over the possible demand occurrences ($m = 0, 1, \ldots, L$):

$$D_{L,t} = \sum_{m=1}^{L} D_{L,t|M=m} P(M = m) \tag{2.1}$$

As $D_{L,t|M=0} = 0$, the case of $m = 0$ is not included in the summation. Expression (2.1) can be conveniently reformulated in terms of inter-arrival times, τ. Given that the forecast is made at time period t, which may or may not coincide with a demand occurrence, we define τ_0 as the time elapsed since the last demand occurrence, and τ_1 as the time until the first upcoming demand occurrence. Successive inter-arrival times are denoted by $\tau_2, \tau_3, \ldots$. Further, given that $m > 0$ demand instances occur during the forecast period, the sum of the corresponding inter-arrival times, A_m, (discarding the elapsed time τ_0) is defined as:

$$A_m = \sum_{k=1}^{m} \tau_k, \quad m = 1, \ldots, L \tag{2.2}$$

Using these inter-arrival times, the probability of having exactly m demand occurrences during the forecast period is obtained as:

$$P(M = m) = \begin{cases} P(A_1 > 0) & \text{if } m = 0 \\ P(A_m \leq L, A_{m+1} > L) & \text{if } m > 0 \end{cases} \tag{2.3}$$

This implies that if demand occurs, the total time of the m inter-arrival times is at least equal to m and at most equal to the length of the forecast period, L. Also, the $(m + 1)$st inter-arrival time extends beyond the forecast period.

We assume that expected demand is strictly additive in the sense that the demand associated with an inter-arrival time is equal to the sum of the expected demand of the spanned time units:

$$D_{\tau_i,t} = \sum_{s=1}^{\tau_i} E(d_{s,t}) \tag{2.4}$$

As a result, the expected demand during the forecast period given $m > 0$ occurrences and the elapsed time before the forecast period, τ_0, can be written as:

$$D_{A_m,t|M=m} = D_{\tau_0} + \sum_{k=1}^{m} D_{\tau_k,t} = D_{\tau_0} + D_{A_m,t} \tag{2.5}$$

The expected total demand during the forecast period (2.1) can be reformulated using (2.3) and (2.5) as a weighted sum of expected demand over the number of demand occurrences m and the sum of intermittent inter-arrival times a, where $m \leq a \leq L$:

$$\begin{aligned} D_{L,t} &= \sum_{m=1}^{L} \sum_{a=m}^{L} D_{a,t|M=m} P(A_m = a, A_{m+1} > L) \\ &= \sum_{a=1}^{L} \sum_{m=1}^{a} D_{a,t|M=m} P(A_m = a, A_{m+1} > L) \\ &= \sum_{a=1}^{L} \sum_{m=1}^{a} D_{a,t|M=m} P(\tau_{m+1} > L - A_m \mid A_m = a) P(A_m = a) \end{aligned} \tag{2.6}$$

where $P(A_m = a, A_{m+1} > L)$ is the probability that the sum of m intermittent inter-arrival times is within the forecast period, while the sum including the $(m + 1)$st demand occurrence is outside this period. The first step in (2.6) changes the summation order between number of occurrences m and total elapsed time a, while the second step redefines the probability of elapsed time a. The final expression conveniently supports the derivation of a closed form expression for the total expected demand in the special case presented next.

2.3.2 Specific application

The properties of the intermittent demand model are illustrated for the case in which the expected demand is proportional to the inter-arrival times: $D_{\tau,t} = \mu\tau$. The inter-arrival time τ follows a geometric distribution, $f(\tau) = p(1-p)^{\tau-1}$, where $\tau = 1, 2, \ldots$ and p is the periodic probability that demand occurs. Note that the definition of inter-arrival times in combination with the assumption of discrete time periods require τ to be at least equal to 1. This distribution is commonly used. It is supported by theory and empirical data, and it conveniently leads to a closed-form solution for the expected total demand during the forecast period (e.g. Eaves and Kingsman, 2004; Syntetos et al., 2012; Willemain et al., 1994). For this special case, equation (2.5) simplifies to:

$$D_{a,t|M=m} = D_{\tau_0} + D_{a,t} = \tau_0\mu + a\mu \tag{2.7}$$

Given these assumptions, we can simplify the expected demand expression (2.6) as:

$$\begin{aligned} D_{L,t} &= \sum_{a=1}^{L}\sum_{m=1}^{a} D_{a,t|M=m} P(\tau_{m+1} > L - A_m \mid A_m = a)P(A_m = a) \\ &= \sum_{a=1}^{L}(\tau_0\mu + a\mu)\sum_{m=1}^{a} P(\tau_{m+1} > L - A_m \mid A_m = a)P(A_m = a) \end{aligned} \tag{2.8}$$

Furthermore, $P(\tau_{m+1} > L - A_m \mid A_m = a)$ refers to the probability that the $(m+1)$st demand occurrence is outside of the forecast period. It is independent of the number of demand occurrences m and equal to $(1-p)^{(L-a)}$, yielding:

$$D_{L,t} = \sum_{a=1}^{L}(\tau_0\mu + a\mu)(1-p)^{L-a}\sum_{m=1}^{a} P(A_m = a) \tag{2.9}$$

Moreover, the sum of the probabilities that the aggregate inter-arrival times, A_m, is equal to a, $1 \leq a \leq m$ is: $\sum_{m=1}^{a} P(A_m = a) = \sum_{m=1}^{a}\binom{a-1}{m-1}(1-p)^{a-m}p^m = p$, the probability of a demand occurrence in any particular period. As a result, the expected demand expression (2.9) can be further simplified to:

$$
\begin{aligned}
D_{L,t} &= \sum_{a=1}^{L} (\tau_0\mu + a\mu)(1-p)^{L-a} p \\
&= \sum_{a=1}^{L} \tau_0\mu(1-p)^{L-a} p + \sum_{a=1}^{L} a\mu(1-p)^{L-a} p \\
&= \tau_0\mu \left[1-(1-p)^L\right] + \mu \left[L - \left(\frac{1-p}{p}\right)(1-(1-p)^L)\right] \\
&= \mu \left[L + \left(\tau_0 - \frac{1-p}{p}\right)(1-(1-p)^L)\right]
\end{aligned}
\tag{2.10}
$$

Expression (2.10) is used to infer the expected demand in following sections.

2.4 Methodology

This section presents the selected forecasting methods, explains the estimation procedures used, and outlines the measures for forecast accuracy and inventory performance to evaluate the methods for multiple data sets.

2.4.1 Data

Five data sets, described in Table 2.1, are used to compare the performance of various methods. For each data set, we include the SKUs that have at least one demand occurrence in the training set. The first data set, hereafter called Electro, is our own and has not been used before. It originates from a global supplier of spare parts for production lines of lamps located at customers' sites. The supplier receives no information about orders, such as whether orders are placed for corrective or preventive maintenance. The data set consists of 1439 SKUs and spans 26 months; see Table 2.1 for the data descriptives. Some of the demand sizes are very large. The other data sets have been used in earlier analyses of the electronics industry, hereafter referenced as ElecInd (Syntetos et al., 2012); Royal Air Force, hereafter referenced as Raf (Syntetos et al., 2009a; Teunter and Duncan, 2009); the automotive industry, hereafter referenced as Auto (Syntetos et al., 2005; Syntetos and Boylan, 2005, 2006); and the US Defense Logistics Agency, hereafter referenced as Navy (Syntetos et al., 2012). See especially Syntetos et al. (2012, Tables 1-4, pp.5-6) and Teunter et al. (2010, Table 1, p.622) for descriptives of these data sets. The ElecInd data set most resembles our own data in terms of the descriptives, also containing various large demand sizes. The Raf data set covers “consumable parts” with no “associated repair activity” (Teunter and Duncan, 2009, p.323).

2.4.2 Selected forecasting methods

The methods discussed in the literature review that are taken into consideration are: single exponential smoothing (SES), Croston’s method, SBA, TSB, and the bootstrap from Willemain et al. (2004). The undamped negative binomial model and hurdle shifted Poisson from Snyder et al. (2012) are excluded, because the numerical optimization is unstable for many of the SKUs in our data sets, which leads to poor results. Two reasons can explain why these methods did not perform well on our data sets: (1) Snyder et al. (2012) use a much stricter selection of SKUs, and (2) their data sets exhibit much lower means and much less variation. They select 1046 out of 2509 SKUs for which the demand data includes at least 10 demand occurrences with at least one in the first 15 and the last 15 months. Our only restriction for the selection of SKUs is that there should be at least one positive demand in the training set (see section 2.4.1). The average demand size for their SKUs is 2.1, which is much lower than the mean demand size for the five data sets we examined: 80.1, 24.1, 14.6, 5.4, and 14.6 (see Table 2.1). The average variance to mean ratio of the demand sizes of their data is 2.3, which is also much lower than most of the average variance to mean ratios of our data sets: 71.0, 31.5, 11.5, 2.4, and 16.0.

The selected methods are compared to our own method, hereafter referred to as DLP, which is derived from (2.10) by smoothing estimates of μ and p as more data becomes available. As our own method is a parametric method, comparing its performance to that of the bootstrap does not directly provide insights into the underlying relation between the elapsed time and the demand size. For this reason, we introduce a variant of the Willemain et al. (2004) bootstrap, hereafter referred to as BootstrapDLP. It is similar to the bootstrap used by Porras and Dekker (2008), but different in that it approximates the probability of a demand occurrence using an empirical distribution of demand occurrences over the lead time that is conditional on the elapsed time. Hence, we can directly compare our DLP method to the other parametric methods, and our BootstrapDLP to the non-parametric methods for fair comparisons, as both are based on the assumption of a positive relation between elapsed time and demand size.

Two-thirds of the available data is used as the training set, and one-third as the holdout sample to compare performance. The parametric methods require initial forecasts and smoothing parameters to generate the forecasts, and these are optimized to minimize the mean square error for the training set (Teunter and Duncan, 2009).

Table 2.1: Descriptives of the empirical datasets used

Data set 1: Electro (26 observations for 1,439 SKUs)									
	Demand per period			Demand sizes			Demand intervals		
	Mean	SD	CV^2	Mean	SD	CV^2	Mean	SD	CV^2
Min.	0.154	0.368	0.207	1.000	0.000	0.000	1.000	0.000	0.000
1st Qu.	2.135	2.916	0.541	4.148	2.970	0.369	1.412	0.784	0.500
Median	6.038	9.100	0.672	11.200	9.533	0.638	1.786	1.193	0.630
Mean	50.170	66.190	0.715	80.100	71.590	0.886	1.942	1.374	0.670
3rd Qu.	25.020	37.030	0.851	42.960	40.370	1.099	2.364	1.806	0.830
Max.	2689.000	4206.000	1.813	7462.000	4372.000	8.883	6.250	5.908	1.601
Data set 2: ElecInd (48 observations for 2,677 SKUs)									
	Demand per period			Demand sizes			Demand intervals		
	Mean	SD	CV^2	Mean	SD	CV^2	Mean	SD	CV^2
Min.	0.104	0.371	0.150	1.091	0.000	0.000	1.000	0.000	0.000
1st Qu.	1.104	2.509	0.362	3.591	3.254	0.593	1.412	0.857	0.548
Median	2.500	5.005	0.534	6.043	6.535	0.894	2.190	1.940	0.781
Mean	18.740	20.400	0.607	24.050	23.260	1.310	2.937	2.663	0.792
3rd Qu.	7.188	11.280	0.767	12.560	14.190	1.437	3.833	3.627	1.010
Max.	5366.000	3858.000	2.252	5366.000	3858.000	15.740	12.000	19.670	2.650
Data set 3: Raf (84 observations for 1,131 SKUs)									
	Demand per period			Demand sizes			Demand intervals		
	Mean	SD	CV^2	Mean	SD	CV^2	Mean	SD	CV^2
Min.	0.060	0.238	0.116	1.000	0.000	0.000	4.150	1.822	0.359
1st Qu.	0.214	0.673	0.252	1.732	1.014	0.235	6.481	5.146	0.690
Median	0.512	1.951	0.289	4.571	3.761	0.508	8.000	6.423	0.801
Mean	1.893	6.923	0.294	14.560	13.680	0.788	8.357	6.694	0.809
3rd Qu.	1.494	5.658	0.336	12.000	11.040	1.000	9.875	8.119	0.925
Max.	51.290	160.400	0.488	307.700	340.300	11.880	16.800	14.440	1.355
Data set 4: Auto (28 observations for 3,000 SKUs)									
	Demand per period			Demand sizes			Demand intervals		
	Mean	SD	CV^2	Mean	SD	CV^2	Mean	SD	CV^2
Min.	0.542	0.504	0.243	1.000	0.000	0.000	1.043	0.208	0.200
1st Qu.	1.458	1.319	0.956	2.050	1.137	0.262	1.095	0.301	0.275
Median	2.333	1.922	1.128	2.886	1.761	0.350	1.263	0.523	0.424
Mean	4.450	3.947	1.173	5.423	3.755	0.436	1.292	0.554	0.410
3rd Qu.	4.167	3.502	1.365	5.000	3.357	0.491	1.412	0.734	0.514
Max.	129.200	122.700	2.558	193.800	101.400	14.070	2.000	1.595	0.997
Data set 5: Navy (60 observations for 3,870 SKUs)									
	Demand per period			Demand sizes			Demand intervals		
	Mean	SD	CV^2	Mean	SD	CV^2	Mean	SD	CV^2
Min.	0.083	0.279	0.130	1.000	0.000	0.000	1.000	0.000	0.000
1st Qu.	0.800	1.882	0.343	3.061	2.498	0.510	1.844	1.462	0.709
Median	2.083	4.148	0.464	5.682	5.423	0.791	2.810	2.746	0.900
Mean	6.696	12.440	0.510	14.590	16.060	1.097	3.646	3.823	0.961
3rd Qu.	5.167	9.988	0.626	12.310	12.840	1.302	4.615	5.203	1.153
Max.	783.900	1219.000	1.703	1327.000	1473.000	12.240	12.000	22.370	2.819

2.4.3 Evaluation

Recent research into intermittent demand examines both forecast accuracy and the impact on actual inventory performance, such as cost, stock volume, and service level to evaluate performance (Eaves and Kingsman, 2004; Snyder et al., 2012; Syntetos and Boylan, 2006; Syntetos et al., 2009a,b; Teunter and Duncan, 2009; Teunter et al., 2010). We also examine both forecast accuracy and inventory performance. Inventory performance is most important as an improvement in forecast accuracy does not necessarily translate into cost savings or a higher service level for inventory management (Syntetos and Boylan, 2006). A simple example is to always forecast zero demand: if the demand is sufficiently intermittent, this is the best performing method in terms of forecast accuracy, but it cannot be used to set inventory levels or for determining order quantities.

Classification by cross-correlation

The inter-arrival time and the squared coefficient of variation of the demand sizes are commonly used to differentiate between particular SKUs (Syntetos and Boylan, 2005, p.12). However, these two quantities do not capture the presence and influence of cross-correlation, so that we require additional dimensions. We propose to determine these cross-correlations using a dynamic structural system where the demand size and the inter-arrival time depend on their lags and on each other, as there is contemporaneous interdependence. The sample of observations $\{d_t\}_{t=1}^{n}$ of the demand for a certain part during the n time periods in the training set can be interpreted as a bivariate sample for the z demand occurrences in the training set $\{s_j, i_j\}_{j\in Z}$, where set $Z = \{1, 2, \dots, z\}$. This sample contains the observations with positive demand together with the time since the previous demand occurrence. Here, we use the notation of Croston's method, so s_j denotes demand size and i_j inter-arrival time for demand occurrence j. The interpretation as a dynamic structural system allows us to estimate the correlations by means of a structural VAR(1) model:

$$\begin{bmatrix} 1 & -\psi_{1,2} \\ -\psi_{2,1} & 1 \end{bmatrix} \begin{bmatrix} s_j \\ i_j \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} s_{j-1} \\ i_{j-1} \end{bmatrix} + \begin{bmatrix} e_{1,j} \\ e_{2,j} \end{bmatrix}$$

For identification of the parameters, the errors in this formulation are assumed to be white noise: that is, normally and independently distributed with covariance matrix I and expectation 0. We can then formulate the reduced-form VAR, impose restrictions, and estimate $\psi_{1,2}$ and $\psi_{1,2}$ using maximum likelihood.

$$\Psi = \begin{bmatrix} 1 & -\psi_{1,2} \\ -\psi_{2,1} & 1 \end{bmatrix}$$

$$\begin{bmatrix} s_j \\ i_j \end{bmatrix} = \Psi^{-1} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \Psi^{-1} \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} \begin{bmatrix} s_{j-1} \\ i_{j-1} \end{bmatrix} + \Psi^{-1} \begin{bmatrix} e_{1,j} \\ e_{2,j} \end{bmatrix}$$

The covariance matrix of the reduced form residuals is: $\Psi^{-1}\Psi^{-1^{\intercal}}$, for which we have an estimate based on the residuals of the reduced-form VAR. The Ψ coefficients provide insight into the cross-correlation between the demand size and the inter-arrival time and can be used to classify SKUs.

SKUs exhibiting strong positive cross-correlation are identified based on the estimated cross-correlation coefficient $\psi_{1,2}$, which after standardizing becomes ρ. We then identify two subsets of SKUs for which there is either a strongly positive cross-correlation, $\rho > 0.5$, or a strongly negative cross-correlation, $\rho < -0.5$, in the training set.

Forecast accuracy

As many observations of intermittent demand are zero, the use of regular forecast accuracy metrics is problematic. Applying conventional forecast accuracy metrics leads to the wrong conclusions for intermittent demand (Teunter and Duncan, 2009). Two metrics have recently been proposed which are now widely adopted to compare accuracy. Hyndman and Koehler (2006) propose to scale the forecast errors based on the in-sample mean average error from the naive forecast, which they call the mean absolute scaled error (MASE):

$$\text{MASE} = \text{mean}(\{\frac{|d_t - \hat{d}_{t|t-1}|}{\frac{1}{n-1}\sum_{t=2}^{n}|d_t - d_{t-1}|}\}_{n+1}^{N})$$

in which n denotes the last time period included in the training set, and N the last time period of available data. If a scaled error is less than one, a method outperforms the naive forecast, and if it is greater than one the method performs worse.

Syntetos and Boylan (2005) advance a different accuracy metric that does not scale the errors, which is the same as the geometric mean absolute error (GMAE):

$$\text{GMAE} = \text{gmean}(\{|d_t - \hat{d}_{t|t-1}|\}_{n+1}^{N})$$

Inventory performance

Inventory performance is examined under often used conditions (Syntetos and Boylan, 2001, 2006; Syntetos et al., 2009a,b, 2012; Teunter et al., 2010). We assume an order-up-to (T, S) policy, where T is a constant review time, assumed to be given, and S is the order-up-to level. The service level is measured as the order fill rate, which is the proportion of demand that can be fulfilled immediately from stock, where the stock position at time t is given by I_t:

$$\text{Service level} = \frac{\sum_{t=n+1}^{N} O_t}{\sum_{t=n+1}^{N} d_t}$$

$$O_t = \min\{d_t, I_t\}$$

Two of the data sets, Electro and Raf, include lead times for each SKU. For the remaining SKUs, we assume a two period lead time. The forecasts from SES, Croston's method, SBA, and TSB are multiplied by the lead time length to give the mean forecast over the lead time. DLP, Bootstrap, and BootstrapDLP directly give forecasts over the lead time. To calculate the order-up-to-level S for a particular target service level, we need more characteristics of the demand distribution over the lead time than just the mean forecast. Although this can be done by simulation (Snyder, 2002), the common approach is to specify a probability distribution and use the demand forecast as the estimated mean, and the smoothed mean square error (MSE) as an estimate for the variance of the distribution of choice. We adopt this approach and use the negative binomial with a smoothing parameter of $\alpha_3 = 0.25$, allowing for non-stationarity of the variance as well as of the mean (Syntetos and Boylan, 2006; Syntetos et al., 2009a,b, 2012; Teunter et al., 2010):

$$\text{MSE}_{t+1|t} = (1 - \alpha_3)\text{MSE}_{t|t-1} + \alpha_3(d_t - \hat{d}_{t|t-1})^2$$

Once we can determine the order-up-to-level S based on the forecast distribution, we can simulate the inventory performance over the holdout sample. This gives us the actual service level attained and the required stock investment necessary according to the forecast distribution. A method that scores higher on service level is not necessarily better than a method which scores lower. We can conclude that one method outperforms another only if both the service level is higher and the necessary stock investment is lower. These results can be easily interpreted with the use of trade-off curves between service level and inventory investment for the various methods (Gardner, 1990), as is also commonly done. In our case the average inventory will be used as a measure for the required stock investment, and the fill rate as a measure of the service level. The results are compared for the various methods over four different service level targets: 85%, 90%, 95%, and 99%. This is very similar to the approach taken by Babai et al. (2014).

Inventory performance is examined here specifically to assess how well the forecasting method can be applied, that is, how well it serves as a basis on which inventory decisions can be made. For this reason, fixed ordering or manufacturing costs are not taken into account.

Financial performance

For the Electro data set, we have price information that can be used to compare the cost benefits of implementing DLP in comparison to the other methods, assuming the same service level target of 95%, for stock investment and lost revenue due to lost sales. As this is sensitive company data, performance is relative to Croston's method, which is chosen because of its popularity. Stock investment is summed over all SKUs as the number of units ordered multiplied by the unit cost. Lost revenue is summed over all SKUs as the number of units that could not be directly fulfilled from stock multiplied by the sales price.

2.5 Results

This section presents the results of the performance of the various examined forecasting methods. Methods are compared based on forecast accuracy, inventory performance, and financial performance for all SKUs or for a subset of SKUs exhibiting strong positive cross-correlation. The five different data sets are discussed in turn. The parametric methods and the bootstrap methods are compared separately, as there often appears to exist a large difference in performance between these two groups.

2.5.1 Forecast accuracy

The forecast accuracy of the forecasting methods is presented in Table 2.2. For both MASE and GMAE, the parametric methods perform much better than the bootstrap methods. SBA is most often the best performing method for both measures, but the difference with some of the other parametric methods, such as TSB, is small. The performance of the DLP method depends on which data set and measure is examined. For the Navy data set, DLP performs worst in terms of MASE, but performs best when GMAE is examined. DLP is actually the best performing method for three out of the five data sets, but only when considering GMAE. For MASE, DLP consistently performs worst of all the parametric methods. A closer examination of the data shows that DLP has a much larger variation in terms of its performance: for some SKUs it performs much better, but for other SKUs much worse. The different scaling of these errors in the accuracy measures leads to different results. Using the cross-correlation to classify the SKUs gives similar results in forecast accuracy performance and offers the same conclusions; these results have therefore not been included here. The overall best performer is SBA, but it remains to be seen whether this advantage translates into actual inventory performance.

Table 2.2: Mean forecast accuracy for the data sets, best cases in bold.

	Electro		ElecInd		Raf		Auto		Navy	
	MASE	GMAE	MASE	GMAE	MASE	GMAE	MASE	GMAE	MASE	GMAE
SES	0.775	33.430	0.834	7.812	1.865	2.507	0.893	2.337	1.796	5.809
Croston	0.835	39.217	0.850	7.950	1.690	2.562	0.897	2.356	1.703	5.787
SBA	0.806	37.341	**0.824**	**7.281**	**1.659**	2.402	**0.880**	**2.278**	**1.678**	5.506
SY	0.824	38.714	0.837	7.875	1.663	2.426	0.895	2.346	1.688	5.632
TSB	**0.744**	32.911	0.831	7.873	1.718	2.325	0.891	2.328	1.700	5.576
DLP	0.950	**20.431**	1.558	9.926	3.990	**2.115**	0.986	2.316	3.417	**4.470**
Bootstrap	0.985	40.968	1.122	11.712	2.195	2.999	1.105	2.779	1.981	6.392
BootstrapDLP	1.133	47.832	1.382	12.288	3.199	4.407	1.103	2.743	2.577	8.027

2.5.2 Inventory performance

Table 2.3 gives an overview per data set of the mean inventory performance for the various methods. For a particular target service level, the inventory policy based on each forecasting method leads to an actual service level attained and a stock investment that was necessary to realize this service level. Table 2.3 lists these results as an average over all SKUs per data set, and for the two subgroups of SKUs which exhibit strongly positive or strongly negative cross-correlation. The differences between target and actually realized service level can be large.

Alternatively, the relative performance of the methods is determined by examining the trade-off curves, which result from combining the data from the different target service levels into one graph per data set. As the graphs become difficult to read with so many methods, and as the performance of many forecasting methods is very similar, we supplement our discussion with specific examples.

For the parametric methods, the inventory performance of SES is worst, and the performance of SBA and TSB is very similar. If we look at the trade-off curves of inventory performance of just DLP and TSB for the Auto data set, we see in Figure 2.1 that DLP outperforms TSB, both for all SKUs and for the group of SKUs that exhibit a strongly positive cross-correlation. TSB outperforms DLP for the

SKUs with a strongly negative cross-correlation, which adheres to our expectation concerning the effect of cross-correlation. The Navy and Raf data sets show similar patterns, but in the ElecInd and Electro data sets, DLP more strongly outperforms TSB, as exemplified in Figure 2.2. The performance difference is largest for the SKUs which exhibit a strongly positive cross-correlation. These results hold if we replace TSB for any of the other parametric methods other than DLP. The performance difference between DLP and TSB is much larger than the difference between TSB and, for example, SBA. We can conclude from these graphs that DLP outperforms the other methods: both over all SKUs and for the SKUs that exhibit strongly positive cross-correlation. Our method not only works well for its intended purposes, but also appears to be quite robust for other cases. The difference in performance between the groups of SKUs demonstrates that our application of the structural VAR, estimated based on the training set, allows us to classify the performance on the holdout sample, so that we are able to determine up front to which SKUs our method should be applied.

Figure 2.1: Inventory performance of DLP and TSB for the Auto data set

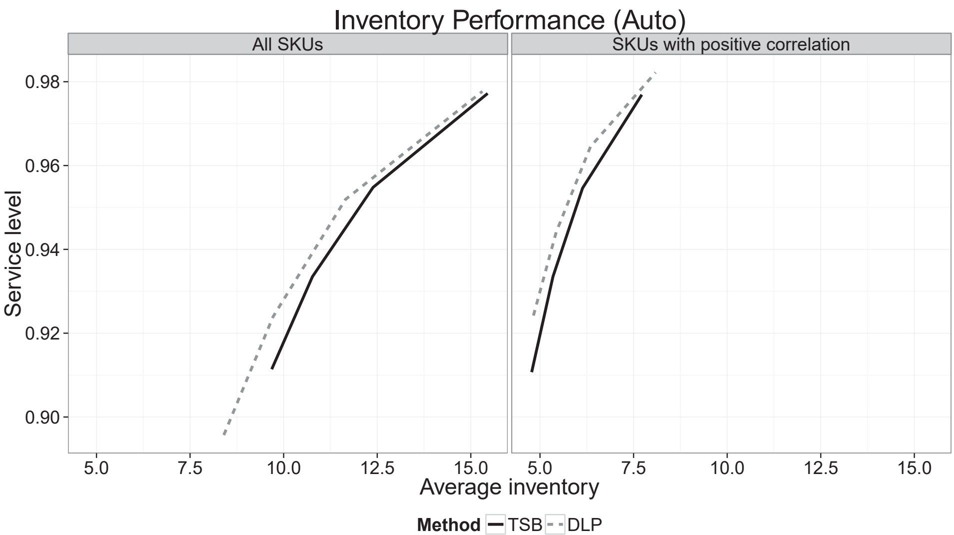


Table 2.3: Mean inventory performance for the data sets for target service level of 95%

Data set 1: Electro (26 observations for 1,439 SKUs)						
	All SKUs		Only positive cross-correlation		Only negative cross-correlation	
	Service level	Stock investment	Service level	Stock investment	Service level	Stock investment
SES	0.916	191.800	0.900	117.300	0.836	121.700
Croston	0.923	199.100	0.896	119.800	0.846	114.100
SBA	0.918	192.200	0.894	116.300	0.838	112.100
SY	0.921	197.800	0.895	118.400	0.843	113.200
TSB	0.911	187.500	0.891	112.100	0.825	113.600
DLP	0.927	185.200	0.940	168.800	0.865	111.000
Bootstrap	0.957	246.400	0.969	166.300	0.915	155.800
BootstrapDLP	0.968	250.300	0.964	166.000	0.930	155.500
Data set 2: ElecInd (48 observations for 2,677 SKUs)						
	All SKUs		Only positive cross-correlation		Only negative cross-correlation	
	Service level	Stock investment	Service level	Stock investment	Service level	Stock investment
SES	0.951	61.860	0.911	18.310	0.961	29.170
Croston	0.956	61.750	0.917	19.560	0.964	27.670
SBA	0.953	59.800	0.916	19.190	0.964	27.230
SY	0.954	61.570	0.916	19.460	0.964	27.460
TSB	0.949	61.190	0.902	17.830	0.957	27.560
DLP	0.971	66.530	0.983	42.900	0.990	31.960
Bootstrap	0.982	79.740	0.970	31.110	0.992	49.440
BootstrapDLP	0.980	83.550	0.953	32.000	0.983	52.020
Data set 3: Raf (84 observations for 1,131 SKUs)						
	All SKUs		Only positive cross-correlation		Only negative cross-correlation	
	Service level	Stock investment	Service level	Stock investment	Service level	Stock investment
SES	0.688	17.070	0.646	9.610	0.708	13.880
Croston	0.725	15.600	0.695	9.034	0.747	12.550
SBA	0.718	15.430	0.690	8.898	0.742	12.430
SY	0.719	15.450	0.692	8.937	0.742	12.440
TSB	0.688	16.200	0.650	9.119	0.707	13.080
DLP	0.889	22.030	0.911	20.240	0.914	18.250
Bootstrap	0.917	34.070	0.910	22.620	0.921	33.870
BootstrapDLP	0.938	41.080	0.939	28.600	0.923	38.540
Data set 4: Auto (28 observations for 3,000 SKUs)						
	All SKUs		Only positive cross-correlation		Only negative cross-correlation	
	Service level	Stock investment	Service level	Stock investment	Service level	Stock investment
SES	0.956	12.490	0.971	5.319	0.956	6.226
Croston	0.955	12.440	0.968	5.278	0.956	6.098
SBA	0.951	12.040	0.965	5.116	0.952	5.873
SY	0.955	12.400	0.968	5.243	0.955	6.076
TSB	0.955	12.390	0.970	5.272	0.955	6.136
DLP	0.952	11.650	0.971	5.380	0.965	6.348
Bootstrap	0.984	15.910	0.995	7.901	0.986	9.074
BootstrapDLP	0.985	15.860	0.997	7.845	0.988	8.978
Data set 5: Navy (60 observations for 3,870 SKUs)						
	All SKUs		Only positive cross-correlation		Only negative cross-correlation	
	Service level	Stock investment	Service level	Stock investment	Service level	Stock investment
SES	0.829	32.170	0.775	34.950	0.778	24.250
Croston	0.843	30.960	0.778	30.930	0.778	22.510
SBA	0.838	30.370	0.775	30.760	0.774	22.070
SY	0.840	30.730	0.775	30.820	0.775	22.290
TSB	0.826	31.210	0.767	33.080	0.769	22.950
DLP	0.913	34.990	0.924	47.740	0.913	31.640
Bootstrap	0.935	47.450	0.902	48.880	0.891	45.220
BootstrapDLP	0.947	52.040	0.909	55.990	0.918	49.870

Figure 2.2: Inventory performance of DLP and TSB for the Electro data set

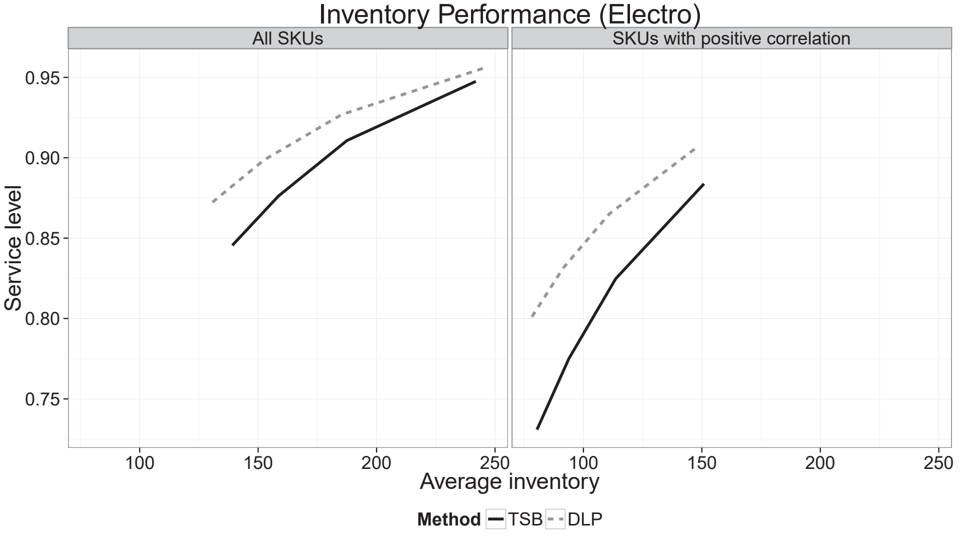


Figure 2.3: Inventory performance of the bootstrap methods for the Auto data set

If we compare the two bootstrap methods, we see two different patterns. For the data sets Auto and Electro in Figure 2.3, BootstrapDLP outperforms Bootstrap, which is especially strong for the SKUs with a strongly positive cross-correlation. For the ElecInd, Navy, and Raf data sets, BootstrapDLP performs better for the lower service level targets, but performs worse than Bootstrap for higher service level targets, see Figure 2.4. In these cases, Bootstrap rapidly improves the service level for a higher stock investment, whereas the gain for a higher investment with BootstrapDLP improves much more slowly. At the lower investment levels, however, BootstrapDLP consistently outperforms Bootstrap.

Disregarding DLP for a moment, Bootstrap consistently outperforms the parametric methods. Though the performance of Bootstrap is better than DLP for the Auto data set, the performance of both is similar for the Electro data set. For the ElecInd, Raf, and Navy data sets, DLP even outperforms Bootstrap as can be seen in Figure 2.5, where the difference is again largest for the SKUs exhibiting cross-correlation.

DLP's superior performance is due to the effect that stock investment is slowly ramped up, instead of consisting of one initial large investment. Figure 2.6 shows how DLP leads to an increase in inventory after there have been no demand occurrences for several periods. This contrasts with methods that assume SKUs can go obsolete at any moment. DLP and TSB both fail to satisfy all demand in Figure 2.6, but DLP is much closer. The demand occurrence at time period 20 leads to an increase in investment for TSB, as the demand occurrence is a clear signal that the SKU is not obsolete. DLP, however, induces the opposite decision of slowly ramping up the inventory again after the demand occurrence, which leads to a much better use of stock investments.

Figure 2.4: Inventory performance of the bootstraps methods for the ElecInd data set

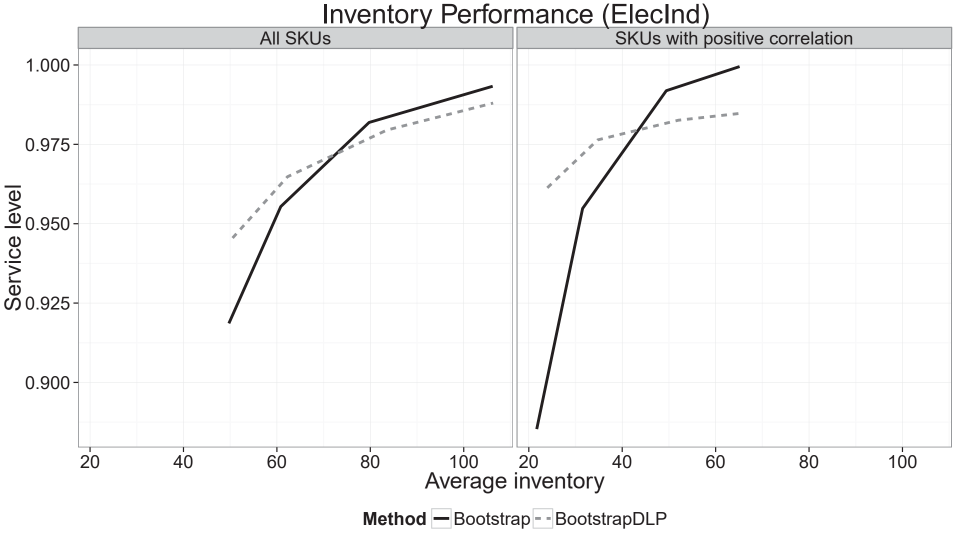


Figure 2.5: Inventory performance of DLP and the bootstrap for the Navy data set

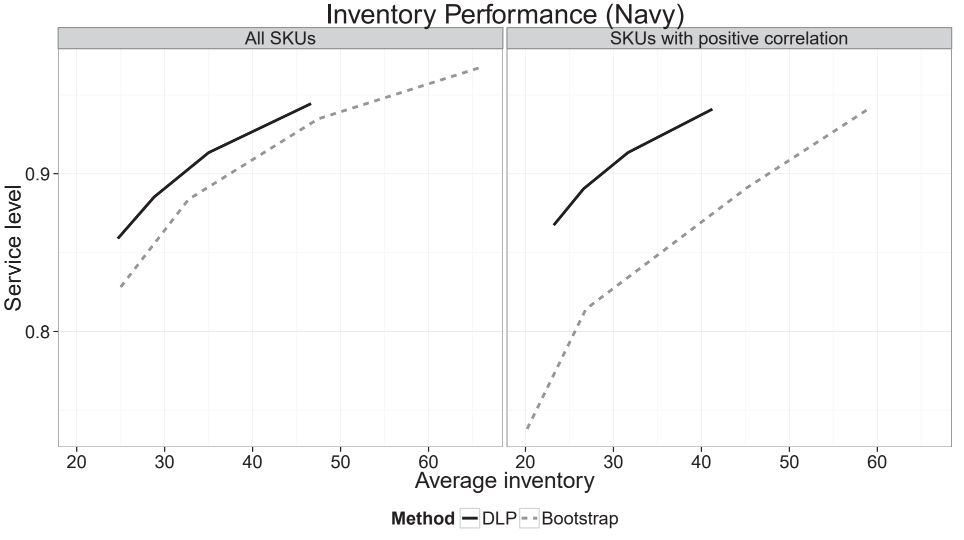
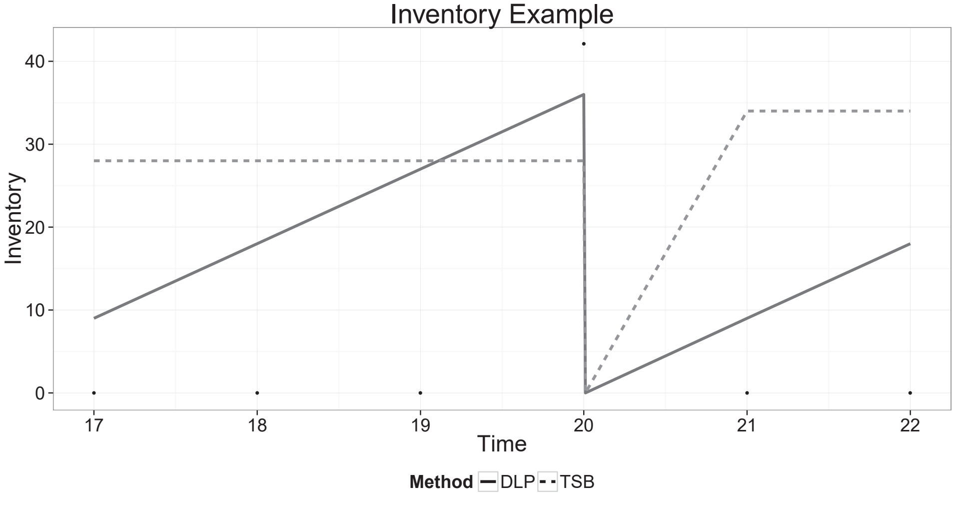


Figure 2.6: This figure shows the inventory, as determined by DLP and TSB, for one SKU over time. The points denote demand; there is one demand occurrence at time 20. DLP and TSB both fail to satisfy all demand, but DLP is much closer. The demand occurrence immediately leads to an increase in investment for TSB, as the demand occurrence is a clear signal that the SKU is not obsolete. DLP, however, slowly ramps up the inventory again, which leads to a much better use of stock investments.



2.5.3 Financial performance

For the Electro data set, the financial performance of the methods can be compared. Table 2.4 lists the quantities of the amount of stock investment necessary and the lost revenue because of stock outs, relative to Croston's method. As Croston is the norm, it scores 1 for each. The lower the investment and lost revenue are, the higher the savings. Table 2.4 gives the results both for all SKUs and for the group with strongly positive cross-correlation. The performance gain is expected in the latter category. However, the performance for all SKUs is quite good. A 4% increase in investment relative to Croston translates into the lowest revenue lost for all methods, with a reduction of 3%. As there is a trade-off between investment and revenue lost there is no method which clearly outperforms the other.

Table 2.4: Financial performance of methods for service level target of 95%: the required stock investment is relative to the investment required for the Croston method, and the lost revenue is relative to the lost revenue of the Croston method. For both quantities the lower number is better. We list the results for all SKUs and for a subset of the data for which the SKUs exhibit positive cross-correlation. Best cases in bold.

	All SKUs		Only positive cross-correlation	
	Investment	Lost	Investment	Lost
SES	0.810	1.000	1.088	0.993
Croston	1.000	1.000	1.000	1.000
SBA	0.950	0.990	0.978	0.990
SY	0.990	1.000	0.991	0.997
TSB	**0.800**	0.990	1.025	0.976
DLP	1.040	**0.970**	**0.863**	**0.963**
Bootstrap	1.010	1.030	1.219	1.069
BootstrapDLP	1.010	1.040	1.234	1.123

For the SKUs that exhibit positive cross-correlation, the superior performance of DLP is substantial. Revenue lost is reduced by 4% which is achieved by investing dramatically less, as investments are reduced by 14%. The better allocation of orders to SKUs translates into a huge gain. There is no trade-off here, but a clear improvement on both counts.

2.6 Discussion and conclusion

We have presented an intermittent demand forecasting method that conditions on elapsed time to anticipate incoming demand and have extensively shown, using five different empirical data sets, that this can substantially reduce both stock investment and lost revenue for spare parts management. We have extensively benchmarked our method against existing forecasting and bootstrapping methods on forecast accuracy and inventory performance, and have shown that its performance is robust under general conditions. Our method is the first to incorporate that activities at the demand side, such as aggregation of demand, preventive and corrective maintenance, can lead to a positive relation between demand size and inter-arrival time of demand occurrences. Our approach has extended the literature by specifically examining the overlooked case of positive cross-correlation between the demand size and the elapsed time, and has shown that substantial financial gains can be realized.

Our method works well for the specific case for which it was designed, that of positive cross-correlation, but, as performance is robust, it can generally be applied to SKUs. Use of our method leads to an increase in inventory performance and financial performance. The classification of SKUs based on a structural VAR analysis shows that not all SKUs should be treated equally, as it successfully led to differentiating between SKUs. The structural VAR can determine for which SKUs our method is most suited, so that the method can be applied to the SKUs for which it will deliver the largest gain in performance.

A general managerial implication of this chapter is that the nature of the demand process is important and has to be considered for forecasting and inventory decisions. Though the risk of obsolescence is important, focusing on this to the exclusion of others can lead to much higher costs, as possible decisions are only taken over a restricted domain. A specific managerial implication of this chapter is that we derive an easy to implement and novel method which can immediately be used for inventory decisions for SKUs, even if context-specific knowledge is unavailable. We also provide the means to assess to which SKUs this method should be applied for the largest gain, so that our method can be used to complement existing the use of existing methods. This also allows managers to apply this method on a smaller scale and facilitates implementation.

The analysis of financial performance shows the importance of applying our method. Our method gave the largest reduction in inventory investment of 14% and even reduced lost revenue by 4%, thus clearly outperforming all other methods. It is easy to estimate and proves to be robust in a range of applications, and is thus generally, and immediately, applicable in practice.

Our specific case was only one implementation of the general method we described. More research is needed to explore the dependency between demand size and elapsed time on empirical data sets, but also to apply more general models, which more broadly incorporate the dynamics between demand size and elapsed time. These dynamics especially come into play due to the product life cycle, so that it can become important to not only classify SKUs once but look and foresee how the characteristics evolve over time, so that they can be incorporated in the method. Our method could also be suited for applications outside of spare parts management.

Chapter 3

Integrated Hierarchical Forecasting

Co-author: J. van Dalen

Abstract

Forecasts are often made at various levels of aggregation of individual products, which combine into groups at higher hierarchical levels. We provide an alternative to the traditional discussion of bottom-up versus top-down forecasting by examining how the hierarchy of products can be exploited when forecasts are generated. Instead of selecting series from parts of the hierarchy for forecasting, we explore using all the series. Moreover, instead of using the hierarchy after the initial forecasts are generated, we consider the hierarchical structure as a defining feature of the data-generating process and use it to instantaneously generate forecasts for all levels of the hierarchy. This integrated approach uses a state space model and the Kalman filter to explicitly incorporate product dependencies, such as complementarity of products and product substitution, which are otherwise ignored. A simulation study, comparing and contrasting existing approaches under varying scenarios of cross-correlations and temporal dependencies, shows the conditions under which an integrated approach is advantageous. An empirical study shows the substantial gain, in terms of forecasting performance as well as inventory performance, of generalizing the bottom-up and top-down forecast approaches to an integrated approach. Specifically, the gains for inventory performance can be as much as a 39% reduction in stock investment. The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of demand forecasting for manufacturers.

Keywords: forecasting, hierarchical, top-down, bottom-up, decision-making.

3.1 Introduction

For organizations, demand forecasting is essential as it drives production, inventory and planning decisions. Demand has to match supply as well as possible to avoid excess inventory and stockouts. Large manufacturers often have SKUs ranging in the thousands, spanning several product categories, each of which requires forecasts. Several decision makers are involved from operations, marketing, sales and finance, who require forecasts at various levels of aggregation. Forecasts are more easily discussed at an aggregated product level, but for production these forecasts have to be available at the SKU level.

SKUs naturally group together in a hierarchy going from the bottom, with individual sales per product, through several intermediary levels, denoting sales for groups of related products at increasingly general aggregation levels, such as product groups and categories, to the top of the hierarchy, which lists total sales. Two commonly used approaches in practice and research start from opposite ends of the hierarchy to generate forecasts for all series: bottom-up forecasting and top-down forecasting (Widiarta et al., 2009). In bottom-up forecasting, base forecasts are generated for product demand at the lowest level in the hierarchy (Gordon et al., 1997). Subsequently, these are aggregated to determine forecasts at higher hierarchical levels. Bottom-up forecasting is commonly contrasted with top-down forecasting, in which forecasts are generated for aggregated demand and disaggregated downwards to determine forecasts at lower levels in the hierarchy (Kahn, 1998). Research stretches over three decades with mixed results as to preference for either bottom-up or top-down forecast approaches.

Both bottom-up and top-down approaches generate forecasts for a selected part of the hierarchy, aggregated upwards or allocated downwards to obtain forecasts for the remaining series. This implies a potential loss of information, as the ignored series can only be recovered under stringent conditions. The loss of information is exacerbated as the selected series are forecasted separately: only after forecasts are generated are they added together, or allocated over items, to generate forecasts for all the series. Thus, product dependencies, such as complementarity of products and product substitution, are explicitly ignored. Yet product dependencies motivate combining similar products in groups and the existence of hierarchies.

Hyndman et al. (2011) introduce a combination approach that uses forecasts of all series in the hierarchy. By taking a linear combination of the bottom-up and top-down forecasts at various hierarchical levels, their approach offers an ensemble of the bottom-up and top-down approaches. The combination entails a post-hoc revision of forecasts to ensure that forecasts add up consistently throughout the hierarchy.

More forecasts are involved than in either the bottom-up and top-down approaches alone, but the initial forecasts are still generated independently.

The bottom-up, top-down and combination approaches use the hierarchy of products only after initial forecasts are generated. By incorporating the hierarchical structure at an earlier stage, i.e. during the generation of forecasts, we introduce an integrated approach. This supersedes the traditional discussion of bottom-up versus top-down forecasting by examining how the hierarchy of products is used in forecasting. This has at least two advantages. First, instead of selecting isolated series for forecasting, all the available data in the hierarchy can be used. Second, product dependencies can be explicitly incorporated, such as complementarity of products and product substitution, while they are otherwise ignored.

A simulation study, comparing and contrasting the approaches from literature under possible cross-correlations and dependencies, shows when an integrated approach is advantageous. An empirical application evaluates the forecasting approaches for one of the largest manufacturers of consumer products, which has hundreds of brands spanning fourteen categories of food products, home and personal care. The empirical study shows a substantial gain, in terms of forecasting performance as well as inventory performance, of generalizing the bottom-up and top-down forecast approaches to an integrated approach.

The remainder of this chapter is organized as follows. In Section 3.2 we present an overview of the relevant literature on hierarchical forecasting and the bottom-up, top-down, and combination approaches for forecasting. We especially focus on the use of the hierarchical structure, product dependencies and demand heteroscedasticity, and we critically evaluate several approaches. In Section 3.3, we outline our simulation and empirical study and introduce an integrated approach for hierarchical forecasting. For the simulation study, we use an optimal forecast as a benchmark to assess forecast performance. For the empirical study, we compare approaches in terms of forecast performance as well as inventory performance and use the company's own forecast as a benchmark. Section 3.4 lists the results and their implications, while Section 3.5 concludes and gives suggestions for future research.

3.2 Theoretical background

Hierarchical forecasting has different forms pertaining to temporal and contemporaneous aspects. Here, we exclusively focus on contemporaneous hierarchies, specifically on products aggregated in groups and categories. This section summarizes the relevant theoretical background on hierarchical forecasting and the approaches of bottom-up, top-down, and the combination approach of Hyndman et al. (2011)

for forecasting. We especially focus on the use of the hierarchical structure, product dependencies and heteroscedasticity in product demand, and critically evaluate approaches.

Over three decades of forecasting literature show mixed results as to a preference for either top-down or bottom-up forecasting. This is not surprising as the performance of the approaches depends on the underlying demand process of products (Lütkepohl, 1984). Due to the additive nature of the hierarchy, with sums of product sales determining group sales, which, in turn, add up to determine category sales, the underlying demand process is transformed at various levels of the hierarchy. Aggregation can lead to substantial information loss, which makes bottom-up forecasting seem favorable (e.g. Edwards and Orcutt, 1969; Orcutt et al., 1968; Zellner, 1969). However, if no important information is lost, benefits can be gained if random noise cancels out (Fliedner, 1999), which makes top-down forecasting seem more favorable. A wide variety of performances is seen, and the nature and extent of differences between top-down and bottom-up are highly dependent upon context (Wei and Abraham, 1981).

Examples of the influence of the demand process are presented by Widiarta et al. (2007), who show that the differences in accuracy of the top-down and bottom-up approaches are only 1% for AR(1) demand processes when the autoregressive coefficient is small. However, for an AR coefficient larger than 1/3, the bottom-up approach is consistently more accurate (Widiarta et al., 2007). Yet, for MA(1) demand processes, performance differences between bottom-up and top-down are negligible (Widiarta et al., 2009).

Dependencies between the demands of different products are a key characteristic of the demand process, and hence a main driver of differences in performance between top-down and bottom-up approaches (Kohn, 1982; Schwarzkopf et al., 1988; Tiao and Guttman, 1980). A particular type of demand dependencies does not unequivocally make either bottom-up or top-down more favorable (Fliedner and Mabert, 1992; Fliedner, 2001; Sohn and Lim, 2007). Stronger negative cross-correlations between individual demand series lead to less variation at an aggregate level, but imply differences between individual product sales. In contrast, stronger positive correlations between individual demand series lead to more variable aggregate sales, but imply that differences at the individual product level are smaller.

This explains why empirical studies are unable to consistently show one approach outperforming the other. Dangerfield and Morris (1992) compare bottom-up and top-down approaches on empirical data and conclude that bottom-up forecasting is more accurate, especially when products are highly correlated. By contrast, Fliedner

(1999) concludes that stronger positive and negative correlations improve the forecast at the aggregate level to such an extent that the top-down approach is more accurate.

An important difference between the bottom-up and top-down approaches is that the latter requires additional measures to allocate an aggregate forecast downwards to lower levels in the hierarchy. Gross and Sohl (1990) compare various ways of determining allocation proportions. A common allocation is based on averaging historical sales proportions, where the unweighted proportion p_j for each product j is determined as its sales y_j relative to the total sales in the product category y over time period T.

$$p_j = \frac{1}{T} \sum_{t=1}^{T} \frac{y_{j,t}}{y_t} \tag{3.1}$$

A common alternative is based on a single, total proportion observed over all time periods, leading to a weighted allocation:

$$p_j = \sum_{t=1}^{T} y_{j,t} / \sum_{t=1}^{T} y_t \tag{3.2}$$

These two allocations perform well in practice (Gross and Sohl, 1990).

The two approaches of top-down and bottom-up can also be combined at intermediary levels in the hierarchy, known as the middle-out approach. Forecasts are generated at a particular level and then aggregated upwards using the bottom-up approach, and allocated downwards using a top-down approach.

Recently, Athanasopoulos et al. (2009) and Hyndman et al. (2011) introduced a different approach, labeled the combination approach, which uses the hierarchical structure to create revised forecasts. This forecasting approach follows two steps: (1) generate independent forecasts for each series in the hierarchy, (2) weight these forecasts according to the hierarchical structure to determine the final forecasts. These final forecasts adhere to the hierarchical structure in the sense that aggregates of the forecasts at the bottom level exactly match forecasts at higher levels in the hierarchy.

The combination approach proposed by Hyndman et al. (2011) is a continuation of earlier work on revising measurements of macro-economic indicators (e.g. Byron, 1978; Solomou and Weale, 1991, 1993, 1996; Stone et al., 1942; Weale, 1985, 1988). A salient difference is that Hyndman et al. (2011) have underlying time series of sales available for each forecast. We introduce notation for hierarchical series to discuss the combination approach, focusing on sales without loss of generality. We have a large vector $\boldsymbol{y}_t$ which contains the n sales series at all levels of the hierarchy. Sales at higher levels are determined by aggregating sales of products at the lowest level

$\boldsymbol{b}_t$. $\boldsymbol{y}_t$ is an $n \times 1$ matrix determined by linear combinations of the $m \times 1$ vector $\boldsymbol{b}_t$ containing sales at the base product level, using an $n \times m$ design matrix $\boldsymbol{S}$ to link sales at each level of the hierarchy with the base level sales:

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{b}_t \tag{3.3}$$

For forecasting, we are interested in the expected $\boldsymbol{y}_t$ and $\boldsymbol{b}_t$. Hyndman et al. (2011) determine the unknown forecasts of product sales, $\hat{\boldsymbol{b}}_t$, as a function of initial forecasts $\hat{\boldsymbol{y}}_t$ generated for each series in the hierarchy by regression, supposing:

$$\hat{\boldsymbol{y}}_t = \boldsymbol{S}\hat{\boldsymbol{b}}_t + \varepsilon_t, \quad \varepsilon_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{3.4}$$

They assume, as a simplifying approximation, that the $n \times 1$ vector $\boldsymbol{\varepsilon}$ is equal to $\boldsymbol{S}$ times a smaller $m \times 1$ vector $\boldsymbol{\varepsilon}$ and use a generalized inverse of the variance-covariance matrix $\boldsymbol{\Sigma}$ to estimate $\hat{\boldsymbol{b}}_t$, which can then be obtained using simple ordinary least squares only involving $\boldsymbol{S}$ and $\hat{\boldsymbol{y}}$ (for details see Hyndman et al., 2011, p.2583). Note that $\hat{\boldsymbol{y}}_t$ contains forecasts for all series, including the product demand forecasts at the lowest level. The revised forecasts are then generated as:

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{S}(\boldsymbol{S}'\boldsymbol{S})^{-1}\boldsymbol{S}'\hat{\boldsymbol{y}}_t \tag{3.5}$$

They conclude that their method is "optimal" because it has "minimum variance amongst all combination forecasts under some simple assumptions" (Hyndman et al., 2011, p.2579). For the multivariate analogue of the Gauss-Markov theorem to apply we require homoscedasticity and absence of cross-correlation in the error terms, requiring the variance-covariance matrix $\boldsymbol{\Sigma}$ to be $\sigma^2\boldsymbol{I}$. For a manufacturer with multiple products, many dependencies may exist among products sales due to complementarity of products and product substitution. Moreover, instead of equal variance of sales over products, it is likely that some series have lower variation and are, as a result, easier to forecast than other series. Series with lower variation give more insight into the demand and so we should weigh the observations of these series higher for forecasting. The restrictions on the variance-covariance matrix $\boldsymbol{\Sigma}$ are easily violated in practice due to possible heteroscedasticity and product dependencies, which can make its proposed estimator highly inefficient, resulting in large mean square forecast errors.

The bottom-up and top-down approaches are based on subsets of all demand series. The bottom-up approach only uses the series at the bottom of the hierarchy as input, while the top-down approach takes the series at the top of the hierarchy as input. These approaches exclusively focus on different parts of the hierarchy, in effect ignoring information. A bottom-up approach cannot benefit from possible noise canceling out at higher hierarchical levels, while the top-down approach suffers from information loss due the use of aggregated series (Fliedner, 1999; Gordon et al., 1997; Kahn, 1998). Both approaches create initial forecasts for the series in the selected parts of the hierarchy only. In the bottom-up approach, forecasts for series at higher hierarchical levels are derived by aggregating the forecasts. In the top-down approach, forecasts for lower hierarchical levels are determined by allocating forecasts downwards in the hierarchy. Moreover, forecasts are generated for each series independently. In contrast, the combination approach can use information from different series more flexibly, as it can use a selection of forecasts generated by both of the other approaches. For example, it can use forecasts at higher levels from a top-down approach, and forecasts at lower levels derived from a bottom-up approach. However, it only uses the hierarchy after forecasts are generated to reconcile forecasts. As the forecasts are generated separately, the hierarchy is not applied to consider the underlying time series of sales for initial forecast generation.

All three approaches ignore the hierarchy when generating the forecasts, and, as a consequence, ignore product dependencies and possible heteroscedasticity in the demand. Exploiting the hierarchy that characterizes the original sales series circumvents the discussion between bottom-up and top-down approaches by directly tackling the underlying demand process.

3.3 Methodology

We employ both a simulation study and an empirical study. The simulation study allows us to compare the forecast performance of the combination approach under different scenarios of product dependencies and heteroscedasticity, in which an optimal integrated approach serves as a benchmark. In the empirical study, we demonstrate the impact, in terms of forecast accuracy as well as inventory performance, of applying the integrated approach to real world sales data from a global supplier in fast-moving consumer goods.

3.3.1 Study 1: Simulation

We assume that the series under observation follow an autoregressive process, subject to hierarchical conditions summarized in $\boldsymbol{S}$, which can be modeled in a multivariate state space form as:

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{S}\boldsymbol{\beta}_t \\
\boldsymbol{\beta}_t &= \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \qquad & \boldsymbol{\varepsilon}_t &\sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}) \\
\boldsymbol{\alpha}_t &= \boldsymbol{\Gamma}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad & \boldsymbol{\eta}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)
\end{aligned} \tag{3.6}$$

The vector $\boldsymbol{y}_t$ consists of sales of products at all levels of the hierarchy. It is derived by adding base sales of products $\boldsymbol{\beta}_t$, as defined by the design matrix $\boldsymbol{S}$. Product sales, $\boldsymbol{\beta}_t$, consist of an underlying state $\boldsymbol{\alpha}_t$ plus measurement noise $\boldsymbol{\varepsilon}_t$. Product state $\boldsymbol{\alpha}_t$ follows an autoregressive process with coefficient matrix $\boldsymbol{\Gamma}$ and disturbances $\boldsymbol{\eta}_t$. The measurement noise of the products at the lowest level in the hierarchy, $\boldsymbol{\varepsilon}_t$, is assumed to be independent, so its variance-covariance matrix is diagonal. The state equation for product sales $\boldsymbol{\alpha}_t$ allows for cross-correlation between the errors, $\boldsymbol{\eta}_t$, represented by the variance-covariance matrix $\boldsymbol{\Sigma}_\eta$.

The design matrix $\boldsymbol{S}$ defines the linear relations between the sales of individual products and their various aggregates. Our simulation considers nine products in three product groups. $\boldsymbol{S}$ is then 13×9, with an identity matrix $\boldsymbol{I}_{9\times 9}$ for the lower nine rows, corresponding with the individual products:

$$\boldsymbol{S}_{13\times 9} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \multicolumn{9}{c}{\boldsymbol{I}_{9\times 9}} \end{bmatrix}$$

The product sales state $\boldsymbol{\alpha}_t$ has an autoregressive structure with coefficient matrix $\boldsymbol{\Gamma}$, following Widiarta et al. (2007) who show that autoregressive coefficients affect forecast performance. As dependencies among products are included through the covariance of errors, $\boldsymbol{\Sigma}_\eta$, $\boldsymbol{\Gamma}$ is a diagonal matrix for all product series. We further restrict $\boldsymbol{\Gamma}$ to a single autoregressive coefficient to limit the number of possibilities and interactions analyzed: $\boldsymbol{\Gamma}_{9\times 9} = \gamma \boldsymbol{I}_{9\times 9}$, where the autoregressive coefficient γ is either a unit root (1), large (5/6), or small (1/6). In the first case, the model is a multivariate local level model, also known as a random walk with noise. The other two cases allow us to examine high and low autoregressive AR(1) processes, respectively; see Table 3.1 for a summary.

We define the variance-covariance matrix $\boldsymbol{\Sigma}_\eta$, representing cross-correlations between and heteroscedasticity of product sales, as block diagonal for the simulation. An unrestricted variance-covariance matrix is impractical for simulating comparable

Table 3.1: **Models employed for simulation**
Overview of three different models used for simulation.

	Autoregressive parameter		Univariate analogue
Model 1	Unit root	$\gamma = 1$	ARIMA$(0, 1, 1)$, local level model
Model 2	Large	$\gamma = 5/6$	AR(1)
Model 3	Small	$\gamma = 1/6$	AR(1)

conditions. Hence, products within the same product group can be correlated, but products that do not belong to the same group have zero cross-correlation. The variance-covariance matrix $\boldsymbol{C}$, a 3×3 matrix reflecting the dependencies among product sales in the same product group, is replicated on the diagonal to have more control over the impact of particular correlation settings:

$$\boldsymbol{\Sigma}_{\eta\ 9\times 9} = \begin{bmatrix} \boldsymbol{C}_{3\times 3} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{C}_{3\times 3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{C}_{3\times 3} \end{bmatrix} \tag{3.7}$$

$$\boldsymbol{C}_{3\times 3} = \begin{bmatrix} 200 & x & y \\ x & 400 & z \\ y & z & 600 \end{bmatrix}$$

We set the measurement noise σ_ε^2 at 200, ensuring that the signal-to-noise ratio is at least one and differs for the three products. Covariances x, y, and z are defined by correlation coefficients, which can have one of the following seven values: -0.75, -0.5, -0.25, 0, 0.25, 0.5, and 0.75. These possibilities lead to 7^3 possible variance-covariance matrices in total, resulting in 7^3 conditions.

The various conditions have to be ranked in order to examine the impact of increasing the magnitude of cross-correlations. This is not straightforward, because of the presence of heteroscedasticity and the simultaneous occurrence of both positive and negative cross-correlation. We take the determinant, $|\boldsymbol{C}|$, also known as the generalized variance, to characterize the magnitude of cross-correlation. More specifically, we use the inverse of the generalized variance, $|\boldsymbol{C}|^{-1}$, so that a larger value signifies higher cross-correlation.

Furthermore, we want to differentiate between an overall positive and negative contribution of cross-correlations, even though both can occur simultaneously. When we aggregate product sales at the product group level, the variance of the aggregate consists of the sum of individual variances and pairwise covariances between products. Positive correlation, or covariance, increases the variance of the sum. Conversely, negative correlation, or covariance, decreases the variance of the sum. We therefore use the sign of the sum of pairwise covariances to differentiate between positive and negative correlation. If $x+y+z > 0$, the variance of the sum increases, characterizing the situation as positive correlation; if $x + y + z < 0$, dependence amongst items decreases the variance of the sum, characterizing the situation as negative correlation.

For each of the three models defined by γ, we have 7^3 conditions defined by $\boldsymbol{C}$, and for each pair of model and condition we perform a hundred iterations of our simulation. Each iteration consists of generating sales for each level in the hierarchy, estimating the forecast approaches on a training set, and calculating forecast performance over a holdout set. Each of these steps will be explained in this order.

For each sales series, we simulate sales data for 1,000 time periods. First, we generate 9×1000 measurement errors ε, which are independently and identically distributed as $N(0, \sigma_\varepsilon^2)$, giving 1,000 errors for each of the nine products. Next, we use the variance-covariance matrix associated with the currently considered condition to simulate 1,000 disturbance vectors from the multivariate normal distribution $N(\mathbf{0}, \Sigma_\eta)$. The autoregressive coefficient is given by the model, so that if we have an initial state, $\boldsymbol{\alpha}_{\mathbf{0}}$, we can recursively compute all time series by first determining $\boldsymbol{\alpha}_t$, then $\boldsymbol{\beta}_t$ and finally $\boldsymbol{y}_t$ for each of the 1000 periods. We initialize each product series using $\boldsymbol{\alpha}_{0,i} \sim N(10000, 600)$. For each iteration the 1,000 time periods are separated into a training sample, consisting of the first 650 observations, and a holdout sample, consisting of the remaining 350 periods.

The training sample generated in each iteration is used to apply the bottom-up approach, the top-down approach, the combination approach, and, for benchmarking purposes, the optimal forecast, all of which will be described next. The training sample of 650 is split into a sample of 450 periods, used to estimate the parameters, and a sample of the remaining 200 periods, which is used to select the forecast methods for evaluation during the holdout sample.

The bottom-up approach generates forecasts at the product level by considering each product series in isolation. No knowledge of the data-generating process is presupposed, and product series are treated as independent. Simple exponential smoothing, Holt's, Holt-Winters, and various ARIMA formulations have been applied to each product series. Parameters are optimized to minimize mean square forecast error over the first 450 periods. The optimal forecast for the product level is included in the methods considered, which is either ARIMA$(0, 1, 1)$, identical to single exponential smoothing, or AR(1). These give the optimal forecasts at the product level for the local level model and autoregressive processes used to simulate the data.

Algorithm 1: Simulation design consisting of models, conditions and iterations to calculate forecast performance

```
Data: Simulation design
foreach three models in γ do
    foreach 7^3 conditions in C do
        foreach 100 iterations do
            Simulate data for 1000 time periods;
            foreach forecasting approach do
                Estimate parameters based on first 450 observations;
                For combination approach, select methods based on
                  performance in next 200 periods;
                For top-down, select allocation rule based on performance in
                  next 200 periods;
                Calculate forecast accuracy using RMSE for each series using
                  holdout set, consisting of final 350 observations;
            end
        end
        Calculate average RMSE per approach and per hierarchical level over
          all iterations;
    end
end
```

The top-down approach generates forecasts at the top level of the hierarchy, and allocates these downwards using (3.1) and (3.2) to create forecasts for the base products. Simple exponential smoothing, Holt's, Holt-Winters, and various ARIMA formulations are applied to the top series, and parameters are optimized to minimize mean square forecast error over the first 450 periods. The optimal method for the top level is included in the methods considered. When we have a unit root in the model, the product sales follow a normal distribution, whose sums are still normally distributed, so that the best forecast at higher levels is, theoretically, provided by single exponential smoothing. When we have autoregressive coefficients other than the unit root, the series at higher levels are still ARMA processes, which are part of the ARIMA formulations (Lütkepohl, 1984). In addition to generating forecasts, the forecasts have to be allocated downwards to the lower levels in the hierarchy, for which we use the two most common ways specified in (3.1) and (3.2), determined using the first 450 periods (Gross and Sohl, 1990).

The third forecasting approach considered is the combination approach of Hyndman et al. (2011). It requires forecasts for each series at each level, which can subsequently be revised into final forecasts using (3.5). The required forecasts in $\hat{\boldsymbol{y}}$ are the best performing forecasts from the bottom-up and top-down approaches. The best performing forecasts are selected by reserving the last 200 periods of the training set as a holdout set, using only the first 450 periods for estimating the methods. For each series, a method is selected which minimizes mean square forecast error for this holdout set. These are then revised to obtain the final forecasts of the combination approach, which is then applied to the actual hold out set.

The last forecast method considered is the optimal forecast, which is an integrated approach and serves as a benchmark. The optimal forecast, as in minimizing mean square error, for our data-generating process is given by jointly forecasting the items at the lowest level, and then aggregating these upwards using $\boldsymbol{S}$ to give forecasts for each series at every level (Engel, 1984; Lütkepohl, 1984). In our case, the optimal forecasts are found by applying the Kalman filter to the state space model as specified by the data-generating process, where parameters are estimated over the training period using maximum likelihood (Durbin and Koopman, 2012; Harvey, 1989). Forecasts for aggregate levels are not generated independently, as in the combination approach, but rather as the sums of product sales forecasts. The Kalman filter traces the forecast errors for each series at each level back to the underlying states, so that the optimal forecasts use information from all series.

After applying these forecast approaches to the generated data within an iteration, we measure forecast accuracy over the holdout sample by calculating the root mean square error (RMSE) of forecast method j for each series i for time periods 651 to 1,000:

$$\mathrm{RMSE}_i^j = \sqrt{\frac{1}{350}\sum_{t=651}^{1000}(\hat{y}_{i,t}^j - y_{i,t})^2} \tag{3.8}$$

where $y_{i,t}$ denotes actual observations and $\hat{y}_{i,t}^j$ the forecast obtained from method j. We take averages of RMSEs for each level in the hierarchy.

In addition to RMSE, we trace the relative performance of the combination approach. Both top-down and bottom-up approaches generate univariate forecasts for parts of the hierarchy. As the combination approach takes a linear combination of forecasts from the top-down and bottom-up approaches it is an ensemble of these approaches. Hence, we expect it to outperform the bottom-up and top-down approaches. Because all of these approaches ignore heteroscedasticity and cross-correlation, and do not use the hierarchy for generating forecasts, they are unable to outperform the optimal forecast. The optimal forecast and the best performing method of the bottom-up and top-down approaches allow us to scale the performance of the combination approach to determine how much it outperforms the bottom-up and top-down approaches and how close it comes to the optimal performance. We scale the performance of the combination approach, $\hat{y}^{\mathrm{ca}}$, using the best forecast method of the top-down and bottom-up approaches, $\hat{y}^{\mathrm{tb}}$, and the optimal forecast, $\hat{y}^{\mathrm{opt}}$, for each series i:

$$\mathrm{sRMSE}_i^{\mathrm{ca}} = (\mathrm{RMSE}_i^{\mathrm{ca}} - \mathrm{RMSE}_i^{\mathrm{opt}})\frac{1}{\mathrm{RMSE}_i^{\mathrm{tb}} - \mathrm{RMSE}_i^{\mathrm{opt}}} \tag{3.9}$$

One simulated iteration results in an RMSE and a scaled performance sRMSE for each series in the hierarchy. We average these values over the hundred iterations performed. Each of these hundred iterations is performed for each combination of one of the three models and one of the 7^3 conditions of possible variance-covariance matrices (see equation (3.7)), resulting in a total of 102,900 simulation runs.

3.3.2 Study 2: Empirical data

As the state space model merely functions as a benchmark for the simulation study, a practical case is used to explore its performance on empirical data. We will demonstrate the impact, in terms of forecast accuracy as well as inventory performance, of applying an integrated approach to real world sales data from a global supplier in fast-moving consumer goods, by comparing its performance to the bottom-up and top-down approaches and the combination approach.

Most of the conventions of the simulation study can be reused. We incorporate all previously used methods and include methods which allow for seasonality, such as seasonal ARIMA and Holt-Winters. For the empirical data, we can no longer assume a data-generating process to define an optimal model, instead we formulate a simple but flexible state space model which incorporates seasonality, allows for market changes, and takes cross-correlation and heteroscedasticity into account.

The data have been made available by one of the world's largest manufacturers of consumer products, which has hundreds of brands spanning fourteen categories of food products, home and personal care. For forecasting, several decision makers are involved from operations, marketing, sales and finance, who require forecasts at various aggregation levels. Forecasts are often discussed at an aggregated product level, but for production these forecasts are transformed to the SKU level.

The forecasting methods offered by the company's IT systems are univariate. As a consequence, the company's forecasting matches the top-down and bottom-up approaches described previously. Forecasts are generated for products at a level called 'forecast unit,' which represents an SKU. A forecast unit can consist of several 'distribution units' to account for small changes in products, such as in artwork or in ingredients. Distribution units are ignored in the forecasting process, as well as in this present study. Similar forecast units group together in so called 'forecast groups.'

Data pertaining to the regular sales of SKUs and product groups, excluding promotions, was obtained from the company. The data includes the statistical forecast generated by the company and the final forecast after judgmental adjustment. This data was collected for the years 2010, 2011 and 2012, yielding 156 weekly time series, which allows us to examine trends and seasonality. The first two years of data constitute the training sample, and the final year serves as the holdout sample. Because of changes to product lines and promotions, historical data of three years is not consistently available for all products. For these reasons, two particular product categories have been selected: foods and personal care.

For foods, we examine two product groups of mayonnaise and ice cream. For personal care, we consider one product group of hair products. These product groups consist of several forecast groups and forecast units. Mayonnaise sauce include three different forecast groups: mayonnaise pots, mayonnaise bottles and mayonnaise with a screw cap. Mayonnaise pots consists of twelve forecast units. Mayonnaise bottles consists of thirteen forecast units. Mayonnaise with a screw cap is excluded, because historical data is incomplete. Ice cream has two forecast groups, labeled jars and cones. Jars consists of 37 forecast units. Cones consists of seventeen forecast units. The hair products products consists of sixteen forecast groups of which one forecast group is analyzed: shampoo, which consists of 27 forecast units. See Table 3.2 for an overview.

Table 3.2: **Product groups at the company**
Overview of forecast groups and forecast units. Total number of forecast units analyzed is 106.

Product category	Product groups	Forecast groups	Forecast units
Foods	Mayonnaise	Mayonnaise pots	12
		Mayonnaise bottles	13
	Ice cream	Jars	37
		Cones	17
Personal care	Hair products	Shampoo	27

By extending the extend model specification (3.6) to more flexibly capture the market for consumer goods, we can apply an integrated approach to this case. Most importantly, we have to include possible trends, because of market changes, and weekly seasonality in the model. At the highest level, product sales $\boldsymbol{\beta}_t$ determine sales at all levels of the hierarchy $\boldsymbol{y}_t$ using the design matrix $\boldsymbol{S}$, which is observed with measurement noise.

$$\boldsymbol{y}_t = \boldsymbol{S}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}) \tag{3.10}$$

The formulation of $\boldsymbol{\beta}_t$ depends upon underlying state variables $\boldsymbol{\alpha}_t$, but is extended to include seasonality δ_t. Though we expect that we can discern one overall seasonal pattern within a product group, the impact of this pattern can be different per product. The seasonal effect can be scaled differently for each product, using a diagonal loading matrix $\boldsymbol{A}$ (Durbin and Koopman, 2012):

$$\begin{aligned} \boldsymbol{\beta}_t &= \boldsymbol{\alpha}_t + \boldsymbol{A}\delta_t \\ \boldsymbol{A} &= \mathrm{diag}(\psi_1, \ldots, \psi_n) \end{aligned} \tag{3.11}$$

We include weekly seasonality in trigonometric form to limit the number of parameters in δ_t to estimate (Durbin and Koopman, 2012). Moreover, rather than having seasonality constant throughout the three year period, we allow seasonality to change over time to allow for market changes. Thus, seasonality is stochastic and its influence can change over the 52 weeks per year. The formulation allows for seasonal effects that are smoothed over time and ensures that the contributions of the seasonal errors ω_{jt} and ω_{jt}^* are not amplified by the trigonometric functions (Proietti, 2000):

$$\begin{aligned} \delta_t &= \sum_{j=1}^{26} \delta_{jt}, \quad s = 52, \quad \lambda_j = \frac{2\pi j}{s} \\ \delta_{j,t} &= \delta_{j,t-1} \cos\lambda_j + \delta_{j,t-1}^* \sin\lambda_j + \omega_{jt}, \qquad \omega_{jt} \sim N(0, \sigma_\omega^2) \\ \delta_{j,t}^* &= -\delta_{j,t-1} \sin\lambda_j + \delta_{j,t-1}^* \cos\lambda_j + \omega_{jt}^*, \qquad \omega_{jt}^* \sim N(0, \sigma_\omega^2) \end{aligned} \tag{3.12}$$

The product sales $\boldsymbol{y}_t$ have independent measurement noise $\boldsymbol{\varepsilon}_t$, as the dependencies between products in a product group is contained in the underlying states $\boldsymbol{\alpha}_t$. The sales per product $\boldsymbol{\alpha}_t$ follow an autoregressive process with diagonal coefficient matrix $\boldsymbol{\Gamma}$ and disturbances $\boldsymbol{\eta}_t$. Possible cross-correlations are in the variance-covariance matrix $\boldsymbol{\Sigma}_\eta$. Sales for products are likely to change over time due to market developments, so that sales can have short-term positive or negative trends. We extend the model to allow for an additive dampened trend $\boldsymbol{\theta}_t$ to be able to incorporate these possible market developments (Durbin and Koopman, 2012):

$$\begin{aligned} \boldsymbol{\alpha}_t &= \boldsymbol{\theta}_t + \boldsymbol{\Gamma}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta) \\ \boldsymbol{\Gamma} &= \mathrm{diag}(\gamma_1, \ldots, \gamma_n) \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim N(\mathbf{0}, \sigma_\zeta^2 \boldsymbol{I}) \end{aligned} \tag{3.13}$$

This model outperformed alternative formulations when applied to the training sample. We apply the model to each forecast group. We use the Kalman filter for forecast error decomposition to efficiently apply maximum likelihood to determine all unknown quantities, including the initial states using the BFGS algorithm, a quasi-Newton method of numerical optimization, with 500 random starts. For n items, we have to estimate $\frac{n^2}{2} + \frac{13n}{2} + 6$ parameters. See Table 3.3 for an overview of the number of estimates needed per forecast group.

Table 3.3: **Parameters and observations per forecast group**
Overview per forecast group of the number of forecast units in the group, the number of parameters that have to be estimated, and the number of observations available in the training period.

Forecast group	Units	Parameters	Observations (training)
Mayonnaise pots	12	156	1,248
Mayonnaise bottles	13	175	1,352
Ice cream jars	37	931	3,848
Ice cream cones	17	261	1,768
Hair shampoo	27	546	2,808

To measure forecast accuracy, we use the mean average percentage error (MAPE), instead of the root mean square error, to obscure the scale of the original series, required in view of the confidentiality of the data:

$$\text{MAPE}_i^j = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{y}_{i,t}^j - y_{i,t}}{y_{i,t}}\right| \tag{3.14}$$

where $y_{i,t}$ denotes actual observations and $\hat{y}_{i,t}^j$ the forecast obtained from method j. We take averages of MAPE for each level in the hierarchy.

In addition to evaluating forecast accuracy, we assess the impact of various methods on inventory performance in terms of stock investment and service levels using an order-up-to (T, S) policy, where T is a constant review time, and S is the order-up-to level. The service level is measured as the sales fill rate, defined as the proportion of demand, d_t, that can be fulfilled immediately from stock, where at time t the stock position is given by I_t, and sales by O_t:

$$\text{Service level} = \frac{\sum_{t=1}^{T} O_t}{\sum_{t=1}^{T} d_t} \tag{3.15}$$

$$O_t = \min\{d_t, I_t\}$$

The service level target is 95%, so that the order-up-to-level S is calculated by obtaining the 95th percentile of the forecast distribution. The state space model gives a forecast distribution, but all other methods only provide point forecasts. By splitting the holdout sample in two, the forecast errors over the first half can be used to approximate the forecast distribution. The mean square error during this period can be used as a measure of variance, which characterizes a fitted distribution, such as the normal distribution. However, this can be restrictive in a practical setting, such as assuming symmetry of forecast errors. An alternative, more flexible forecast distribution is given by bootstrapping the forecast errors, using a hundred draws with replacement, to derive an empirical forecast distribution from which the 95th percentile can be obtained. We will use the bootstrap for all methods, including the state space model, to derive the 95th percentile to ensure that methods are comparable. We then simulate inventory levels over time, and determine inventory performance over the remaining holdout sample by calculating the actual service level attained.

Average inventory is used as a proxy for the required stock investment. We include the company's own forecast in this evaluation, and scale the average inventory for each method using the average inventory needed when we use the company's own forecast.

3.4 Results

We evaluate the performance of the forecasting approaches in a simulation study, and assess the potential of our integrated approach in an empirical application. The simulation study allows us to determine the impact of heteroscedasticity and dependencies on forecast performance for the combination approach, which will appear to be extensive, and shows when an integrated approach is advantageous. The empirical application demonstrates that forecast accuracy and inventory performance can be substantially improved with an integrated approach to forecasting hierarchical series, which balances possible information loss and gain at the hierarchical levels.

Table 3.4: **Forecast accuracy (RMSE) for three simulation models**
This table shows the forecast accuracy, in terms of RMSE, of the optimal forecast, the combination approach, and the best performing bottom-up and top-down approach for the three different models, the three boundary conditions, and the three hierarchical levels of top, middle and base.

	Negative correlation			Independent			Positive correlation		
	Top	Middle	Base	Top	Middle	Base	Top	Middle	Base
Local level model ($\gamma = 1$)									
Optimal forecast	21.90	12.18	26.05	67.67	38.97	26.58	104.02	59.72	26.02
Combination approach	31.01	17.24	36.88	95.75	55.14	37.61	147.83	84.87	36.98
Bottom-up/top-down	33.31	18.52	39.62	140.98	81.18	55.37	154.41	88.65	38.62
AR(1), large ($\gamma = 5/6$)									
Optimal forecast	21.90	12.18	26.05	67.67	38.97	26.58	104.02	59.72	26.02
Combination approach	30.86	17.16	36.71	95.90	55.22	37.67	146.65	84.19	36.68
Bottom-up/top-down	31.92	17.75	37.97	126.97	73.11	49.87	168.41	96.69	42.12
AR(1), small ($\gamma = 1/6$)									
Optimal forecast	21.64	13.02	26.37	66.92	38.39	26.25	106.37	62.89	26.52
Combination approach	30.69	18.47	37.40	94.70	54.33	37.15	150.36	88.90	37.49
Bottom-up/top-down	33.47	20.13	40.78	125.32	71.90	49.16	168.97	99.89	42.13

3.4.1 Study 1: Simulation

The simulation study evaluates forecast accuracy, measured as RMSE (see equation (3.8)), of the bottom-up, top-down, combination, and optimal approach for each of the 7^3 different conditions of each of the three models: the local level model and the strongly and weakly autoregressive processes. The 7^3 conditions capture different types of product dependencies, through the variance-covariance matrix of errors in product sales (see equation (3.7)). The conditions are grouped based on the sign of the sum of pairwise covariances to determine whether correlations are overall positive or negative amongst products, labeled positive and negative correlation respectively. The remaining condition, in which the variance-covariance matrix is diagonal, is labeled independent. For the groups of positive and negative correlation, conditions are ranked based on the inverse of the generalized variance to differentiate between the extent of dependencies.

Due to the large number of conditions, we restrict our attention to only one condition for each of the three groups. By selecting the independent condition and the conditions with the largest inverse generalized variance for the groups of positive and negative correlation, we derive three boundary conditions, ranging from the condition most strongly characterized by negative correlation to the condition most strongly characterized by positive correlation. Table 3.4 summarizes the forecast accuracy, averaged over a hundred iterations, for the boundary conditions for each of the three models and the three hierarchical levels. The table combines the bottom-up and top-down approach and lists the best performance of the two. Comparing the forecast accuracy of each forecasting approach for a particular condition over the three models reveals that performance differences between models are small, implying that results are robust to the model specification chosen here. For the optimal forecast, the largest difference in accuracy is 3.17 between the RMSE for the local level model and the autoregressive process at the middle level, which is a difference of 5%. For the combination approach, the largest difference of 4.71 is between the RMSE for the two autoregressive processes at the middle level, which is also 5%. For the best performing bottom-up and top-down approach, the largest difference is 15.66, which is 11%. Overall, for around 80% of the comparisons, the approaches differ only slightly in terms of performance, between 0 and 1%, over the three models.

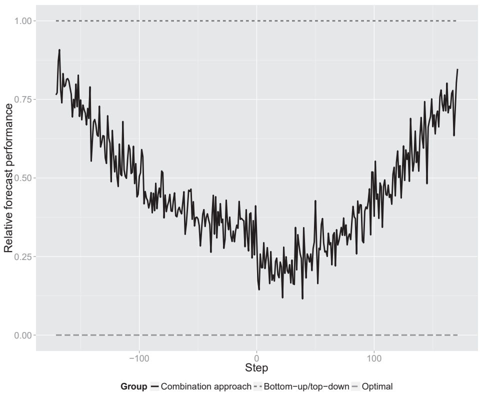
These differences are small compared to the substantial differences, ranging from 50% to 80% for the top and middle level, between conditions for a particular model and forecasting approach. For instance, note the dramatic increase of the RMSE for the optimal (from 21.90 to 104.02), combination (from 31.01 to 147.83) and the best performing bottom-up and top-down approach (from 33.31 to 154.41) for the top level of the local level model, moving from the boundary condition of negative correlation to the boundary condition of positive correlation. These huge performance differences demonstrate the impact of positive and negative cross-correlation.

The combination approach is consistently more accurate than the best performing bottom-up and top-down approach. For the negative and positive boundary conditions, the increase in forecast accuracy ranges from 3%, a difference of 1.06 for the negative boundary condition for the autoregressive process with large coefficient, to 13%, a difference of 21.76 for the positive boundary condition. Improvements are largest for the independent condition, ranging from 24%, for the autoregressive processes, to 32% for the local level model.

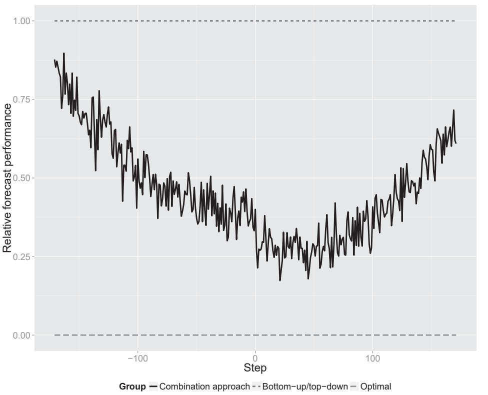
Yet, the combination approach is consistently less accurate than the optimal forecast. In the negative and positive boundary conditions, the difference between the combination approach and the best performing bottom-up and top-down is always substantially less than the difference between the combination approach and the optimal forecast. For instance, in the case of the local level model and negative boundary condition, the RMSE of the combination approach for the top series is 31.01, which is an improvement over the RMSE of the bottom-up and top-down approach of 33.31, but is far removed from the RMSE of 21.90 of the optimal forecast. For the independent condition, the difference between the combination approach and the optimal forecast is much less.

The forecast accuracy of the combination approach is consistently between that of the optimal forecast and the bottom-up and top-down approach. This allows one to scale the RMSE of the combination approach using Equation (3.9) to trace the relative performance of the combination approach in terms of the optimal forecast and the best performing bottom-up and top-down approach. By simplifying the forecast performance of the approaches to a single scaled number in each condition, we trace the forecast performance in more detail over all conditions, and extend the discussion beyond the boundary conditions. Figures 3.2a, 3.2b, and 3.2c depict the average scaled forecast performance for the local level model, the autoregressive process with a large coefficient, and the autoregressive process with a small coefficient, respectively. On the y-axis, the figures show the performance of the combination approach, scaled between the forecast performance of the bottom-up and top-down approaches and the optimal forecast. The 7^3 conditions are ordered on the x-axis, with the independent condition in the middle at zero, conditions from the negative group placed to the left of the middle, and conditions from the positive group positioned to the right of the middle. Moving further away from the middle corresponds with conditions having an increasingly higher inverse generalized variance. The boundary conditions, previously discussed, are placed at the lowest, highest and zero value on the x-axis. Performance is similar for the three models, although variability in performance decreases as the autoregressive coefficient becomes smaller.

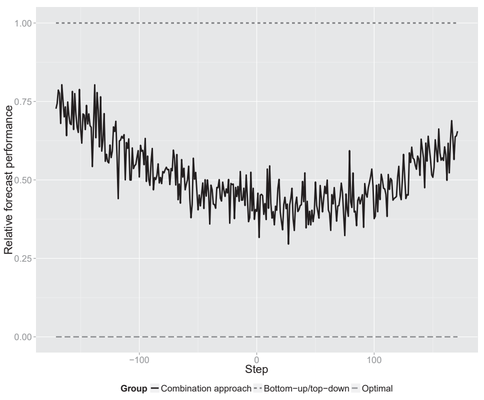
Figure 3.1: **Scaled forecast accuracy**
This graph shows the relative forecast performance of the combination approach. The conditions are ordered on the x-axis with the negative boundary condition at the left, the independent condition in the middle, and the positive condition at the right. Conditions are ordered in between these conditions based on group and inverse generalized variance. The gap in the middle between the combination approach and the optimal approach shows the impact of heteroscedasticity, and the widening gap to the sides shows how correlations impair the forecast accuracy of the combination approach.



(a) Local level model ($\gamma = 1$)



(b) AR(1), large ($\gamma = 5/6$)



(c) AR(1), small ($\gamma = 1/6$)

In the middle of the graphs, and slightly to their right, the performance of the combination approach is closest to the performance of the optimal forecast with a difference of around 10% to 20%, and the bottom-up and top-down approaches are most strongly outperformed. For the independent condition, the observed gap between the performance of the combination approach and the optimal forecast is ascribed to heteroscedasticity. The combination approach suffers from inefficiency, due to ignoring differences in the signal-to-noise ratios of series. It weighs all forecasts for series equally, whereas some forecasts are better, because they are forecasts for series with smaller variation. As correlations increase between products, the forecast accuracy of the combination approach degrades, and moves further away from the optimal forecast.

3.4.2 Study 2: Empirical data

By comparing and contrasting the various approaches under possible forms of cross-correlations and dependencies, the simulation study shows that an integrated approach can outperform the other approaches. The gain does not necessarily translate to practice, as in practice the data-generating process is not defined and presupposed. We therefore present an empirical application of real world sales data from a global supplier in fast-moving consumer goods to further demonstrate the performance of an integrated approach. The performance gain appears substantial in terms of forecast accuracy as well as inventory performance.

Table 3.5 shows the forecast accuracy of the integrated approach, the combination approach, and the best performing bottom-up and top-down approaches for the five forecasting groups. The results show that the forecast performance of the integrated approach is substantially and consistently higher than that of the other approaches. The integrated approach dominates the other approaches in terms of forecast accuracy over all forecast groups. Compared to the best performing bottom-up and top-down approaches, the integrated approach leads to an improvement in forecast accuracy of between 26%, for ice cream jars, and 51%, for ice cream cones. Though the integrated approach has its worst performance for ice cream cones, its MAPE is less than half the MAPE of the bottom-up and top-down approaches.

The performance of the combination approach appears to be unstable, as it does not persistently outperform the best performing bottom-up and top-down approach in all forecast groups. In the case of ice cream cones, its MAPE is worse than the MAPE of the bottom-up/top-down approach. In all other cases, its improvement in forecast accuracy over the bottom-up and top-down approaches ranges between 4%, for mayonnaise bottles, to 22% for hair shampoo. Overall, the combination

Table 3.5: **Forecast accuracy (MAPE) for empirical data**
This table shows the forecast accuracy, in terms of MAPE, of the integrated approach, the combination approach, and the best performing bottom-up and top-down approach for each of the five forecast groups.

	Foods				Personal care
	Mayonnaise		Ice cream		Hair
	Pots	Bottles	Jars	Cones	Shampoo
Integrated approach	39.54%	38.20%	34.93%	41.20%	29.11%
Combination approach	48.08%	61.79%	41.95%	91.10%	38.57%
Bottom-up/top-down	61.04%	64.66%	46.92%	83.37%	49.63%

approach constitutes an improvement over the best performing bottom-up and top-down approach, but is in turn outperformed by the integrated approach.

The superior forecast performance of the integrated approach translates into substantial financial savings for inventory management. Table 3.6 summarizes inventory performance, as measured by achieved service level (see equation (3.15)) and required stock investment relative to the company, of the three approaches and the company's own forecast. For all approaches, the realized service levels are lower than the target of 95%, which means that all approaches underestimate the variation in sales. In the cases of ice cream jars and shampoo, not a single approach is able to achieve a service level higher than 89%. The actual service level achieved using the integrated approach is much closer to the target of 95% than all other methods. For ice cream cones the integrated approach achieves a service level of 92.34%, after which the best performing approach achieves a service level of only 78.29%. Hair shampoo is similar, though the difference is somewhat smaller, with the integrated approach achieving a service level of 88.72%, after which the best performing approach only achieves a service level of 79.75%. Mayonnaise pots is an exception, where achieved service levels of all approaches are high and above 90%. The more accurate point forecasts of the integrated approach allow a better approximation of the 95th percentile of the forecast distribution, resulting in higher service levels for all other forecast groups.

Stock investment for each approach is relative to the stock investment required by the company. The company is consistently outperformed in all product groups. All approaches are almost consistently achieving higher service levels with much lower stock investments. Switching to the best performing bottom-up and top-down approach entails better performance in all product groups, with the exception of mayonnaise bottles. The outcome for bottles cannot be directly compared, because though the bottom-up and top-down approach means a 5% decrease in stock investments, the service level also drops by 3 percentage points. For all other groups, the bottom-up and top-down approach substantially improves performance, especially for ice cream jars, where stock investment is lowered by almost 8%, and service level is even increased by almost 15 percentage points. Mayonnaise pots is another example, where the bottom-up and top-down approaches constitute a similar reduction in stock investment of almost 8%, and service level is increased by 10 percentage points.

The inventory performance of the combination and the bottom-up and top-down approaches cannot be directly compared in the cases of ice cream jars and cones. In these cases, the bottom-up and top-down approaches achieve a higher service level, but does so at the expense of a higher stock investment. The combination and bottom-up and top-down approaches have similar performance overall. In the case of mayonnaise pots, the bottom-up and top-down approach performs better, but in the cases of mayonnaise bottles and hair shampoo the combination approach performs marginally better.

The integrated approach results in a dramatic drop in the required stock investment, reducing the stock investment needed, based on the company's current forecast, by one-third in all product groups. The biggest reduction is in ice cream jars, where the integrated approach reduces the required stock investment by 39%. The largest reduction offered by another approach is given by the combination approach, also for ice cream cones, which only entails a reduction of 11%. The integrated approach gives its smallest stock investment reduction for hair shampoo, which is still equal to 27%. Compared to the other approaches, the integrated approach offers substantial gains.

Table 3.6: **Inventory performance**
This table shows the inventory performance of the three approaches and the company's own forecast based on a service level target of 95%. Stock investment for each approach is relative to the stock investment needed based on the company's own forecasts. All numbers are percentages.

	Foods								Personal care	
	Mayonnaise				Ice cream				Hair	
	Pots		Bottles		Jars		Cones		Shampoo	
	Stock	Service	Stock	Service	Stock	Service	Stock	Service	Stock	Service
Integrated approach	67.88	91.01	69.37	91.65	61.24	88.63	64.01	92.34	72.54	88.72
Combination	95.44	93.37	93.41	91.76	89.16	80.57	88.72	72.68	91.89	79.75
Bottom-up/top-down	92.10	94.41	94.52	74.09	92.11	84.40	90.49	78.25	92.73	77.88
Company's own	100	84.27	100	77.70	100	69.97	100	78.29	100	71.31

3.5 Discussion and conclusion

We introduced an integrated hierarchical forecasting approach to forecast the demand of products at different, but hierarchically-related aggregation levels. The approach supersedes the traditional comparison of bottom-up and top-down approaches (Fliedner, 1999; Kahn, 1998), by generating forecasts at all hierarchical levels and incorporating all available information, rather than only using selected parts of available data. The integrated approach avoids ex-post revising of forecasts, as is done in the combination approach (Hyndman et al., 2011), as generated forecasts are already reconciled and respect the additive restrictions placed on the series by the hierarchy.

Our simulation study, which compares and contrasts existing approaches under possible cross-correlations and dependencies, demonstrates under which conditions our integrated approach is advantageous. Furthermore, our empirical study shows the substantial gain, in terms of forecasting performance as well as inventory performance, of generalizing the bottom-up and top-down forecast approaches to an integrated approach. All available information is used, product dependencies are taken into account, such as the complementarity of products and product substitution, and other features of the series are incorporated as well, such as seasonality, which are otherwise ignored.

The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of forecasting for manufacturers. Even overlapping groups of products can be easily accommodated. The large reductions in stock investments, up to as much as a 39%, show that the forecast performance directly translates to large financial gains, and is highly relevant for forecasting processes at companies. The advantages of formulating the integrated approach as a state space model are that outliers, missing values, and extra information, such as pertaining to promotions, can be easily, and flexibly, included (Durbin and Koopman, 2012; Harvey, 1989). The results of the simulation study and empirical study show that future research has to broaden its scope beyond the bottom-up and top-down approaches, as these approaches are too restrictive, by ignoring dependencies and only using parts of the available data, which comes at serious financial costs.

Chapter 4

Chasers, Smoothers and Departmental Biases: Heterogeneity in Judgmental Forecasting

Co-authors: J. van Dalen and L. Rook

Abstract

Judgmental forecasting has gained considerable research attention, leading to detailed knowledge on how biases and heuristics hamper corporate forecasting. Research so far has primarily studied judgmental forecasting using aggregate measures over large groups of individuals, overlooking the likely differences between groups or between individuals. This is unfortunate, because forecasting heterogeneity - i.e., individual differences in forecasting behavior - exists, and complicates drawing conclusions based on aggregate results. In the present study, we find confirmation of this claim, and specifically for the existence of two distinct forecaster types: one characterized by overreaction to forecast errors (labeled chasers); the other characterized by underreaction to forecast errors (labeled smoothers). Extending the models used in earlier behavioral experiments, our approach relies on wavelets and state space modeling to incorporate forecasting heterogeneity. We demonstrate that contextual biases can only be meaningfully explored after controlling for the forecaster's inclination towards chasing or smoothing. We further show that departmental biases persistently impact judgmental forecasting, even if forecasts are constructed to be free of

intentional biases. Our findings have important repercussions for theory building based on evidence derived from aggregate results, but also have practical relevance for training and hiring of forecasters, and orchestrating forecasting processes in companies.
Keywords: forecasting, heterogeneity, biases, decision-making, incentives.

4.1 Introduction

This chapter studies the phenomenon of judgmental forecasting, a vital component of the corporate forecasting process, which greatly affects corporate supply chain performance (Fildes et al., 2008; Syntetos et al., 2011, 2010). Judgment has been labeled an "indispensable component" of forecasting, because judgmental forecasting is an important and widely conducted activity in organizational practice. It is often used to capitalize on valuable tacit or domain-specific knowledge that is not captured by models (Fildes et al., 2008). However, it introduces various biases inherent to human decision-making, leading to suboptimal decision-making (Lawrence et al., 2006). To compensate for these suboptimal consequences, researchers have in recent years begun to study decision-making from a behavioral perspective (Gino and Pisano, 2008).

The upcoming field of behavioral operations management has, among others, documented how well-known general biases such as the confirmation bias (the tendency of people to only find and use information that is consistent with their own ideas), conservatism (the tendency of people not to adjust their beliefs when they receive new information), overconfidence (the tendency of people to put too much weight on their own judgment), and illusion of control (the tendency of people to believe they control or influence an outcome that they demonstrably have no influence over) may lead to suboptimal decisions (Gino and Pisano, 2008). The seminal study of Schweitzer and Cachon (2000) into decision-making in a newsvendor experiment emphasizes how two specific heuristics – demand chasing, and anchoring and adjustment – influence decision making in sequential judgmental decision-making. The first, demand chasing, refers to the widely observed phenomenon that decision makers in a newsvendor experiment are strongly affected by the last observed demand (Bolton and Katok, 2008; Bostian et al., 2008; Schweitzer and Cachon, 2000). The second, anchoring and adjustment, occurs when forecasters partly adjust, or smooth, their forecasts in reaction to forecast errors (Goodwin and Wright, 1993; Hogarth and Makridakis, 1981).

A major problem of the current knowledge on judgmental biases and the performance of judgmental forecasting is that most of the evidence on is at an aggregate level, encompassing large groups of individuals (in the case of experiments based on the newsvendor model see e.g. Bolton and Katok, 2008; Bostian et al., 2008; Kremer et al., 2011; Schweitzer and Cachon, 2000). This is problematic, because it overlooks the existence and impact of forecasting heterogeneity, which refers to the possibility that forecasting behavior differs systematically between individuals. It may well be the case that two types of forecasters differ in the extent to which they overreact or underreact to forecasting errors, and display chasing or smoothing behavior. Such heterogeneity of individual biases possibly leads to inaccurate aggregate results, which do not reflect individual behavior (Lau et al., 2014).

Moreover, judgmental forecasting in corporations is often a group activity rather than an individual activity. The generation of demand forecasts in large organizations typically requires coordination among different departments, such as sales, operations, and finance, often embedded in a Sales and Operations Planning (S&OP) process. Case study evidence in that respect seems to suggest that group forecasts in corporate setting often are undermined by opposing interests that are played out in S&OP negotiations via the exchange of intentionally inflated or deflated forecasts Nauta and Sanders (2001); Oliva and Watson (2009, 2011). This interplay of unintentional, introduced by biases and heuristics in decision-making, and intentional, influenced by the setting, biases has been largely ignored so far in the experimental studies on judgmental forecasting.

The contribution of this chapter, therefore, lies in the assessment of the consequences of heterogeneity for judgmental forecasting. We demonstrate, using an approach relying on wavelets and state space modeling, that forecasting behavior indeed differs systematically between individuals. That is, forecasters can be divided into people who overreact to forecast errors and display chasing behavior, and people who underreact to forecast errors, and thus display smoothing behavior. This observation has important repercussions for the assessment of departmental roles and incentives of the forecaster in a group setting, and especially for orchestrating forecasting processes in companies.

The remainder of this chapter is organized as follows. In Section 4.2 we give an overview of the relevant literature about judgmental forecasting. In Section 4.3, we outline the proposed method to extend the earlier analyses. In Section 4.4, we describe the set-up of our behavioral experiment and how the data was collected. Section 4.5 lists the results and their implications, while Section 4.6 concludes and gives suggestions for future research.

4.2 Theoretical background

This section gives an overview of the literature on judgmental forecasting and examines various approaches used to analyze forecasting behavior observed during experiments. Our own approach builds on the models from earlier work by Kremer et al. (2011) and Bostian et al. (2008).

4.2.1 Judgmental forecasting

Performance of judgmental forecasting depends on the characteristics of the series, the source and nature of information, and the presentation of the task (Gönül et al., 2009; Lawrence et al., 2006; Moritz et al., 2014). Performance also relies on the behavior of forecasters (Fildes et al., 2009; Syntetos et al., 2009b), and their training (McCarthy Byrne et al., 2011). Much research has focused on eliciting forecast biases (Goodwin and Fildes, 1999; O'Connor et al., 2000; Massey and Wu, 2005). Bias and inefficiency in judgmental forecasts can be so strong as to "mask any contribution of contextual information to accuracy" (Lawrence et al., 2000, p. 161), possibly due to information overload and anchoring. Moreover, these biases and limitations seem to be persistent, as learning effects of forecasters appear to be limited, and forecasters are unwilling to admit to mistakes and revise their forecasts (Kirchgässner and Müller, 2006; Syntetos et al., 2009b).

4.2.2 Individual biases: forecasting heterogeneity

The call that "research should be conducted which [...] fully recognises the importance of the individual" (Goodwin et al., 2007, p. 392) has inspired a vast research on the issue if, and to what extent, individual forecasters are affected by behavioral predispositions and/or biases of various nature, and how these affect behavior and performance. De Véricourt et al. (2013), for instance, examine how differences in gender and attitudes towards risk can explain variations in forecast performance. Moritz et al. (2013, 2014) and Cantor and Macdonald (2009) demonstrate how psychological differences determine the way people perform in judgmental forecasting. Typically, these studies aim to explain variations in forecast performance in terms of particular a priori traits rather than differences in forecasting behavior per se.

This is unfortunate, because it is possible to trace forecast behavior itself. Single exponential smoothing, a popular forecast method, can be viewed as a reflection of human behavior in the form of an anchor and adjustment model (Lawrence and O'Connor, 1992). The method generates forecasts by anchoring on the last forecast and adding an adjustment based on the last forecast error:

$$\hat{d}_{t+1|t} = \hat{d}_{t|t-1} + \alpha(d_t - \hat{d}_{t|t-1}) \tag{4.1}$$

where $\hat{d}_{t+1|t}$ denotes the demand forecast for period $t+1$ made at time t, and d_t denotes the observed demand at time t. The smoothing parameter α can be viewed as a behavioral component in the anchor and adjustment model to capture the individual's reaction to forecast errors. Moreover, in this approach, single exponential smoothing is a proxy for an individual's trial-and-error learning. Sterman (1989) shows that this anchor and adjustment model explains subjects' behavior well, while Schweitzer and Cachon (2000) in their seminal paper report evidence for such an anchor and adjustment model in the context of a newsvendor problem.

Kremer et al. (2011) examine forecasting behavior using this anchor and adjustment model in an experiment in which participants have to forecast a demand series d_t generated by a local level model, also known as a random walk with noise:

$$\begin{aligned} d_t &= l_t + \varepsilon_t, \qquad & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ l_t &= l_{t-1} + \nu_t, \qquad & \nu_t &\sim N(0, \sigma_\nu^2) \end{aligned} \tag{4.2}$$

where σ_ε^2 and σ_ν^2 change between various conditions in their experiment. Single exponential smoothing is optimal for a local level model (Durbin and Koopman, 2012), so that the intuitive anchor and adjustment method is optimal if participants weigh forecast errors correctly. The signal-to-noise ratio $q = \sigma_\nu^2/\sigma_\varepsilon^2$ determines the smoothing parameter α^* that minimizes the mean squared forecast error. This parameter can be thought of as the steady state of the Kalman gain, the optimal weighing factor for new information, when the Kalman filter is applied (Durbin and Koopman, 2012):

$$\alpha^* = \frac{\sqrt{q(q+4)} - q}{2} \tag{4.3}$$

where $0 \leq \alpha^* \leq 1$, because variances σ_ν^2 and σ_ε^2 are non-negative.

In this setup, Kremer et al. (2011) are able to compare participants' forecast adjustments to the optimal smoothing value. In addition, they generalize the exponential smoothing model (4.1)to capture participants' forecasting behavior as a random walk with noise:

$$\begin{aligned} \hat{d}_{t+1|t} &= \hat{l}_{t+1|t} + \hat{r}_{t+1|t} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \\ \hat{l}_{t+1|t} &= \theta \hat{d}_{t|t-1} + \alpha(d_t - \hat{d}_{t|t-1}) + (1-\theta)A \\ \hat{r}_{t+1|t} &= \hat{r}_{t|t-1} + \beta(\hat{l}_{t+1|t} - \hat{l}_{t|t-1} - \hat{r}_{t|t-1}) \end{aligned} \tag{4.4}$$

The model corresponds to double exponential smoothing in $\hat{l}_{t+1|t}$ and $\hat{r}_{t+1|t}$, as a random walk with noise can give the impression of short-term trends. To further generalize the anchor and adjustment model, θ allows for anchoring on either the previous forecast or a fixed long-term value A.

Motivated by the unobservable $\hat{r}_{t+1|t}$, Kremer et al. (2011) estimate the parameters of model (4.4) through the following specification:

$$\begin{aligned} \hat{d}_{t+1|t} = a_0 + a_1(d_t - \hat{d}_{t|t-1}) + a_2\hat{d}_{t|t-1} + a_3(d_t - d_{t-1}) \\ + a_4(d_{t-1} - d_{t-2}) + a_5(\hat{d}_{t-1|t-2} - \hat{d}_{t-2|t-3}) + \eta_t \end{aligned} \tag{4.5}$$

A drawback of this approach, however, is that it introduces identification problems. Model (4.5) is not a special case of model (4.4), but is a distinct model. The parameters of model (4.4) are not uniquely identified in terms of those of model (4.5). Kremer et al. (2011) refrain from parameter restrictions, and instead choose a transformation for each variable of interest, which contradicts the generalized model (4.4).

Based on their estimated parameter α, they conclude that forecasters overreact to forecast errors in relatively stable environments, but underreact to errors in relatively unstable environments. This is a conclusion of no minor importance, but in two conditions of relatively unstable environments, however, their results are not quite straightforward. Specifically, Kremer et al. (2011) report an α of 0.68 (s.e. 0.04) and 0.56 (s.e. 0.04) in two conditions which do not significantly differ from the optimal α of 0.61. Moreover, the average α's from their descriptive measures exceed 0.7 in these cases, which is much higher than the optimal α. These findings do not support their overall conclusion that in a relatively unstable environment forecasters underreact (Kremer et al., 2011), and invite further research into characterizing the effect of a relatively unstable environment.

The anchor and adjustment heuristic has been observed in experiments with decision makers facing independent and identical draws from a stationary distribution, even when explicitly told and instructed that the draws are independent. This behavioral tendency has become well-known as the demand-chasing heuristic (Bolton and Katok, 2008; Schweitzer and Cachon, 2000). Bostian et al. (2008) examine this heuristic with participants facing independently and uniformly distributed demand, and include autoregressive dynamics to examine learning effects:

$$\begin{aligned} \hat{d}_{t+1|t} &= \hat{d}_{t|t-1} + \alpha_t(d_t - \hat{d}_{t|t-1}) \\ \alpha_t &= (1 - \Delta_\alpha)\alpha_{t-1} \end{aligned} \tag{4.6}$$

Because demand is serially independent in their experiment, the optimal α is zero. The constant Δ_α can be interpreted as a stepwise proportional decrease of the bias. The model implies that the existing bias decreases linearly over time. From the estimates, Bostian et al. (2008) conclude that experience improves performance, and that the bias linearly decreases over time. However, estimating learning as a linear effect imposes serious restrictions on participants' behavior, because the behavior of forecasters can be nonlinear and complex (Trapero et al., 2011).

Moreover, inferences based on sample averages and standard deviations can be misleading when applied to behavioral heterogeneity, especially when heterogeneity is used to imply multi-modality in behavioral patterns (Juran and Schruben, 2004). The question therefore arises how individual heterogeneity influences aggregate results (Su, 2008, p.586). As aggregate data does not adequately describe the population of individual decision makers when their behavior is highly heterogeneous (Lau et al., 2014), it may well be the case that some of the conclusions drawn by previous studies may be misleading.

4.2.3 Departmental biases

It is common practice in corporate forecasting processes that demand forecasting is a result of interdepartmental decision-making. Unfortunately, however, the various departments within the company – such as operations, sales, finance and marketing – may also, at least partially, have opposing interests (Nauta et al., 2001, 2002). Forecasts can be influenced by managerial deliberations other than achieving forecast accuracy (Syntetos et al., 2009b). For instance, the forecast may be colored by organizational goals causing the forecast to be intentionally biased (Lawrence et al., 2000) – i.e., the "result of deliberate and rational decision making behavior on the part of the forecasters" (Lawrence and O'Connor, 2005, p.3). The bias can arise and be consistent with rationality because of asymmetric loss functions across forecasters (Aretz et al., 2011; Ashiya, 2009), or because forecasters can intentionally inflate the forecast to ensure that suppliers give them priority (Syntetos et al., 2009b; Terwiesch et al., 2005).

Evidence exists that interdepartmental forecasts are influenced by the various, sometimes conflicting, incentive schemes and agendas between departments (Oliva and Watson, 2009, 2011; Yaniv, 2011). Kuo and Liang (2004) show that forecasters may also be affected by their departmental roles even when there are no incentives in place. This implies that the departmental role itself suffices to trigger different behavior. This observation is supported by Önkal et al. (2012), who show that assigning varying roles to members of a group, even without incentives, has a significant effect on the forecasts made by the group. When members are given the role of forecasting executive, marketing director or production director, they are less satisfied with consensus forecasts and display a strong commitment to their own roles when compared to members without a particular role (Önkal et al., 2012). It is, therefore, important – also for corporate, interdepartmental, forecasting processes that: "[w]e must always remember that forecasts are rarely, in themselves, disinterested and innocent products of the group process in which they are produced and this reality should cause us to reconsider the way in which we evaluate forecasts." Wright and Rowe (2011, p. 12).

4.3 Proposed method

In studying forecasting heterogeneity, we follow suggestions previously made by Lau et al. (2014) and Su (2008) to explicitly elicit individual behaviors, away from approaches that focus on aggregate results. Starting from Kremer et al. (2011) and Bostian et al. (2008), we propose a method to examine forecasting behavior as well as individual and departmental biases that allow for behavioral heterogeneity in the forecasting process.

Similar to Kremer et al. (2011), we employ the local level model (4.2) to simulate demand for a judgmental forecasting experiment. We extend models (4.5) and (4.6) of Kremer et al. (2011) and Bostian et al. (2008) with an alternative based on wavelets and state space modeling to overcome previous limitations and analyze both individual and departmental biases while accounting for heterogeneity.

4.3.1 Heterogeneity in judgmental forecasting

A formal characteristic of forecasts based on subsequent demand is that they form time series. Performance analyses of such forecasts commonly use aggregate measures, such as forecast accuracy or coefficients of estimated regression models, to assess, for instance, the bias. When these forecasts are produced by judgmental forecasters working with the same demand series, the resulting individual differences are typically taken into account by modeling the multivariate series as panel data with random effects, in which the estimated effects are interpreted based on means and standard deviations; see especially Bostian et al. (2008) and Kremer et al. (2011). This approach is implicitly driven by the assumption that a single true value exists for each model parameter, around which participants are randomly located, and that that these aggregate estimates represent actual individual behavior.

There is, however, no a priori reason to assume that behavioral patterns are at all times symmetrically distributed around a common value. Heterogeneity could well implicate the existence of distinct types of forecasting behavior that are associated with different parameter values. Lau et al. (2014) demonstrate that relying on means leads to estimated behavior that does not reflect the behavior of any of the participants. For instance, model (4.4) allows a weighting θ between the previous demand forecast and a fixed long-term value as an anchor, but aggregate results will be misleading if θ has distinct values representing different types of behavior.

In our approach, instead of lumping together forecast series of all participants in an experiment, we first determine if groups of participants can be identified with similar forecasting behavior. In order to identify such groups, we cannot rely on straightforward clustering of forecast series given that this ignores the time structure of forecasts: forecasts are dependent on past values, and the time structure cannot simply be ignored (Chaovalit et al., 2011). For clustering, we need a limited number of independent dimensions, which can be achieved by transforming the forecasts before clustering (Gavrilov et al., 2000; Lin et al., 2004).

Even though the Fourier transform is a common method of transforming time series, it is not suitable for our analysis. The Fourier transform projects the original time series onto several sinusoidal functions, each corresponding to a particular frequency component (Hamilton, 1994). Unfortunately, this projection captures information in the frequency domain, but not in the time domain, because the sinusoidal functions are not localized in time and continue indefinitely. Information from the time domain can only be recovered under certain conditions (Hamilton, 1994). A sufficient condition for preventing loss of information is that the examined series is stationary. But in many cases, the series to forecast is non-stationary, as the mean and other moments of the underlying process can depend upon time. As people exhibit nonlinear and complex behavior (Trapero et al., 2011), judgmentally forecasting non-stationary series is unlikely to result in series of forecasts that meet the strict condition of stationarity. Even the windowed Fourier transform, which puts the sinusoidal functions in a window localized in time (Hamilton, 1994), still loses most information from the time domain, as the condition of stationarity is still imposed within each window.

Wavelets offer an alternative transformation preserving information from both the time and frequency domain (Gençay et al., 2001). Wavelets can flexibly represent a wide array of time series and do not require the time series to be stationary; see the introductions by Percival and Walden (2006) and Struzik (2001). Essentially, a wavelet is a zero-mean function with finite oscillations that fade out. The wavelet transformation decomposes a function into a set of wavelets.

A set of infinite wavelets is equal to $L^2(\mathbb{R})$, the space of measurable functions that are square integrable (Struzik, 2001). Hence, we can decompose any function $x(t)$ in $L^2(\mathbb{R})$, for which $\int |x(t)|^2\, dt < \infty$, into wavelets. This ensures that $x(t)$ has finite oscillations and is localized in time (Percival and Walden, 2006). We represent the function $x(t)$ as a series of successive approximations, based on linear combinations of wavelet basis functions, $\psi_{m,n}$:

$$\begin{aligned} x(t) &= \sum_{m,n} c_{m,n} \psi_{m,n}(t) \\ c_{m,n} &= \langle x, \psi_{m,n} \rangle \end{aligned} \tag{4.7}$$

The wavelet ψ is parametrized in terms of time (or location) by n and in terms of dilation (or scale) by m. Unlike the Fourier transformation, the wavelet basis thus captures both location and scale (Abramovich et al., 2000).

We are interested in transforming a discrete time series x. Daubechies wavelets, a family of wavelets defining a discrete wavelet transform, are commonly used for discrete series (Crowley, 2007). Daubechies wavelets have no closed-form expression and are specified by a single parameter p as D(p) (Gençay et al., 2001), which determines the number of vanishing moments of the approximation (Ogden, 1997). Large values of p allow for representations of higher degree polynomials. A D(6) can have constant, linear and quadratic signal components, while a D(8) can include cubic signal components in addition to the components of polynomials of lower degree. A D(8) with eight coefficients is a common choice in financial and economic applications and is therefore adopted here (Struzik, 2001). A small p does not seem applicable in the present context due to possibly nonlinear and complex behavior of forecasters (Trapero et al., 2011). For each individual time series, the eight orthonormal coefficients for the wavelet transformation are found using numerical integration to evaluate the inner product in (4.7); see Percival and Walden (2006).

Transforming the time series of forecasts using the D(8) discrete wavelet transformation captures the time series in a small number of independent dimensions (Chaovalit et al., 2011). These eight wavelet coefficients not only incorporate how strongly forecasters react to forecast errors, but also capture the value of their last forecast, thus incorporating both the behavioral component α and the anchor from the anchor and adjustment model (4.1).

If forecasts of all n participants are transformed, we have eight coefficients per participant in an $n \times 8$ matrix. This matrix is used to determine if groups of forecasters with distinct types of forecasting behavior exist. By means of k-means++, an adjustment to k-means (Arthur and Vassilvitskii, 2007), clustering involves choosing

k centers minimizing the within-cluster sum of squares, the Euclidean distance, of datapoints y using 500 random starts:

$$\underset{C}{\operatorname{argmin}} \sum_{y} \min_{c \in C} \|y - c\|^2$$

Various values of k will be examined and a value will be chosen based on the ratio of the between-cluster variance to the total variance.

4.3.2 Generalized forecasting model

The generalized forecasting model (4.4) proposed by Kremer et al. (2011) allows for both the perception of short-term trends and the anchoring on either the last forecast or on a fixed long-term value. Instead of estimating this model in reduced form, which introduces problems of identification, we estimate model (4.4) by treating the unobserved quantities as latent variables in a state space model. We employ maximum likelihood, derived using the Kalman filter to decompose the prediction error (Durbin and Koopman, 2012), to calculate the unknown quantities, such as the distributions of the errors, the initial states, and the parameters of interest ($\alpha,\ \beta,\ \theta,\ A$), using the BFGS algorithm, which is a quasi-Newton method of numerical optimization, with 500 random starts.

We modify the formulation of generalized model (4.4) to reflect the use of panel data, where the index i differentiates between participants. Moreover, we have random effects over time in ξ_t, and remaining individual disturbances in η_{it}:

$$\begin{aligned}
\hat{d}_{i,t+1|t} &= \hat{l}_{i,t+1|t} + \hat{r}_{i,t+1|t} + \xi_t + \eta_{it}, \quad \xi_t \sim N(0, \sigma_\xi^2), \quad \eta_{it} \sim N(0, \sigma_\eta^2) \\
\hat{l}_{i,t+1|t} &= \theta \hat{d}_{i,t|t-1} + \alpha(d_t - \hat{d}_{i,t|t-1}) + (1-\theta)A \\
\hat{r}_{i,t+1|t} &= \hat{r}_{i,t|t-1} + \beta(\hat{l}_{i,t+1|t} - \hat{l}_{i,t|t-1} - \hat{r}_{i,t|t-1})
\end{aligned} \tag{4.8}$$

Indicator variables can be included in the equation for $\hat{d}_{i,t+1|t}$ to estimate the effects of various experimental conditions on the one-period ahead forecast.

We further extend this state space model to accommodate learning effects. Bostian et al. (2008) explore a linearly decreasing bias in model (4.6). More flexibility can be attained by making the parameter of interest time-varying. Specifically, we formulate the adjustment α_{t+1} as a random walk, to determine if there is learning towards the optimal behavior and explore how this learning changes over time:

$$\alpha_{t+1} = \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon) \tag{4.9}$$

The parameter values can be derived at each time period by using the Kalman smoother to determine the mean of the parameter conditional on all demand forecasts in the sample (Durbin and Koopman, 2012).

Summarizing, consistent with Kremer et al. (2011), we employ the local level model (4.2) to simulate demand for our judgmental forecasting experiment. We extend models (4.5) and (4.6) of Kremer et al. (2011) and Bostian et al. (2008) with an alternative based on wavelets and state space modeling to overcome previous limitations and analyze both individual and departmental biases, while accounting for heterogeneity. If different types of forecasting behavior using wavelets and clustering are observed, we can estimate our model on subsets of forecaster data to explore the consequences of heterogeneity.

4.4 Experimental design and data

In our experiment, participants are invited to take on the role of a company forecaster. After studying a time series of 18 periods of historic demand, participants provide a forecast for the next period. They then see the actual outcome and their forecast accuracy. Participants iterate through these steps 18 times, making forecasts on a one-period ahead rolling window. The forecast of the expected demand is neutral in the sense that participants privately forecast the demand for the next period and separately propose a desired production quantity. This proposed production quantity is then shared with another manager, and forms the basis for determining the production quantity, explained below. The setup is such that the forecast should be free from intentional biases.

We employ a two-by-two-by-two experimental design, varying with respect to departmental role, incentive scheme, and the behavior of the other manager. That is, participants are randomly assigned the role of either an operations manager or a sales manager. The operations department is focused on production and inventory levels, which may lead forecasters to deflate their proposed production quantities. The sales department is concerned with sufficient product availability so that there are no lost sales, which may lead forecasters to intentionally inflate their proposed production quantities. In terms of incentives, participants are either penalized for outcomes straying from their department's objective, which is either minimizing obsolescence or lost sales, depending on the role, or for outcomes straying from the company's objective, which is maximizing profit by minimizing ex post inventory error.

The production quantity in each time period is determined as the average of two separate inputs, one offered by the participant, the other obtained from a computer agent which represents the other manager. The computer agent can have a neutral, sales or operations role—that is, if the agent is not neutral, its role is complementary to the participant's role, meaning that the computer agent takes on the role of an operations manager when the participant is a sales manager, and vice versa. The computer agent uses single exponential smoothing and the Kalman filter to forecast the demand distribution of the next period, and proposes as production quantity either the 50th, 33th or 66th percentile of the estimated forecast distribution, depending on whether it has a neutral, operations, or sales role, respectively.

Table 4.1: **Experimental data over the four conditions**
The four experimental conditions are based on the two roles of operations and sales, and the incentives of either department or company.

	Department incentive		Company incentive	
	Operations	Sales	Operations	Sales
Total: 357	85 (24%)	89 (25%)	92 (26%)	91 (25%)

Analyses are based on 357 participants (240 men and 117 women with an average age of 21) who are randomly allocated over the different conditions. The number of participants for each role and incentive is listed in Table 4.1. Students of a Business Administration program participated as part of their coursework. They were familiar with the topic. Control questions were used to check whether the participant remembered their role and incentive scheme at the end of the experiment, and whether they understood the forecasting task in the experiment. Respondents who did not answer the control questions correctly were removed from the analyses. Respondents who made typographical errors during the experiment were also left out. As forecasting rounds are dependent, a simple input error influences subsequent rounds and we cannot simply correct obvious errors, interpolate or treat particular inputs as missing values. Out of the initial 467 participants (321 men and 146 women with an average age of 21), 110 participants (24%) are dropped due to typographical errors, which corresponds to an input accuracy of over 99%. No other selection criteria were used.

Behavioral experiments are commonly conducted with students to ensure that analyses are based on a large number of participants (e.g. Bolton and Katok, 2008; Bostian et al., 2008; Kremer et al., 2011; Schweitzer and Cachon, 2000), which is not problematic given that experienced managers and students usually exhibit the same behavior (Bolton et al., 2012). Nevertheless, we have replicated the experiment with 72 professional forecasting and/or demand planners from various manufacturing companies. Even though the analysis cannot be as extensive as with the student sample, the obtained data is used to replicate our previously introduced wavelets and state space modeling approach with a group of practitioners as a robustness check.

Similar to Kremer et al. (2011), we use local level model (4.2) to simulate the demand. We simulate the necessary initial condition using $l_0 \sim N(500, \sigma_\nu^2)$. The variances σ_ε^2 and σ_ν^2 are set to 100, so that our simulated demand closely resembles conditions three and six of the experiments of Kremer et al. (2011). A single demand series is generated and used for each participant. The optimal alpha can be determined using (4.3) and is approximately 0.618.

For forecast accuracy, we calculate the root mean square error (RMSE) per participant i for the forecasting part, which starts at $t = h$ and ends at $t = n$, as follows:

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{n-h+1} \sum_{t=h}^{n} (\hat{d}_{i,t|t-1} - d_t)^2} \tag{4.10}$$

As described above, in our experiment, we have 18 time periods of historic demand, and participants must provide forecasts for 18 periods, so that $h = 19$ and $n = 36$.

We calculate the overall forecast bias per participant as:

$$\mathrm{Bias}_i = \sum_{t=h}^{n} (\hat{d}_{i,t|t-1} - d_t) \tag{4.11}$$

$$\mathrm{RelativeBias}_i = \mathrm{Bias}_i \bigg/ \frac{\sum_{t=h}^{n} d_t}{n-h+1} \tag{4.12}$$

To analyze the evaluation of biases over time, an additional measure is needed throughout the rounds of the experiment. Examining differences between forecasts and demand in each round means that differences are dependent on the specific demand outcome, due to the variability of disturbances. As these differences depend on the original demand series, we also include a performance measure that removes the influence of the original series as much as possible. By subtracting the optimal forecast in each time period, $\hat{d}_{t|t-1}$, rather than the actual demand, we can derive the bias relative to the optimal forecast within the sample for each time period t, allowing us to trace the forecast bias over time:

$$\mathrm{BiasBench}_{i,t} = \hat{d}_{i,t|t-1} - \hat{d}_{t|t-1} \tag{4.13}$$

4.5 Results

In the following section, we first examine if and to what extent wavelets capture the original forecast series, and whether we can distinguish between distinct types of behavior using k-means++ clustering. As anticipated, differentiating between behaviors leads to a partitioning of the participants from the experiment. We estimate the generalized forecasting models for different groups of participants and show the influence of individual behavior and departmental biases by considering heterogeneity, roles, incentives and learning.

4.5.1 Forecasting heterogeneity

We transform each series of forecasts using the discrete wavelet transform D(8), giving a matrix containing the eight coefficients for each of the 357 participants. The transformations capture the original series well. That is, applying the inverse wavelet transform on the coefficients and comparing these to the original series yields an average RMSE of 11.21.

The participants are clustered using k-means++ with the coefficients of the D(8) transformation. To determine the number of centers k, we successively apply k-means++ with k ranging from 2 to 8. A k of 2 has a high ratio of between-cluster variance to total variance of 78%. Increasing k marginally affects the ratio in small steps. A k of eight gives a ratio of 91%. Additional centers affect the larger of the two centers when k is two, but do not substantially improve the variance explained. We, therefore, examine the case of two forecast groups in the remainder of this section.

Table 4.2 summarizes the allocation of roles and incentives within the observed two groups of forecasters. Group 1 consists of 271 participants, representing 76% of all respondents, and Group 2 consists of 86 participants, corresponding to 24% of all respondents. The respondents in the two groups are approximately uniformly distributed over the various roles and incentives. Of particular relevance to the present discussion, the two centers appear independent of conditions in the experiment ($\chi^2_{3,0.05} = 0.6994, p = 0.8734$), and thus seem to capture individual forecasting behavior instead.

Figure 4.1 depicts the eighteen subsequent forecasts for each group in addition to the time series of demand. Compared to Group 2, Group 1 produces a more volatile forecast series, implying that forecasters in Group 1 more strongly adjust their forecast as a reaction to forecast errors than Group 2. Group 1 tends to display demand chasing behavior—these ‘chasers‘ undervalue their own forecast, and even overreact to forecast errors. Admittedly, in time periods 26 and 27 of Figure 4.1, chasers stay close to the just observed demand, but in the preceding time period 24,

Table 4.2: **Distribution of respondents in each group over the various conditions of roles and incentives**
The two groups found with clustering are approximately uniformly distributed over the various roles and incentives.

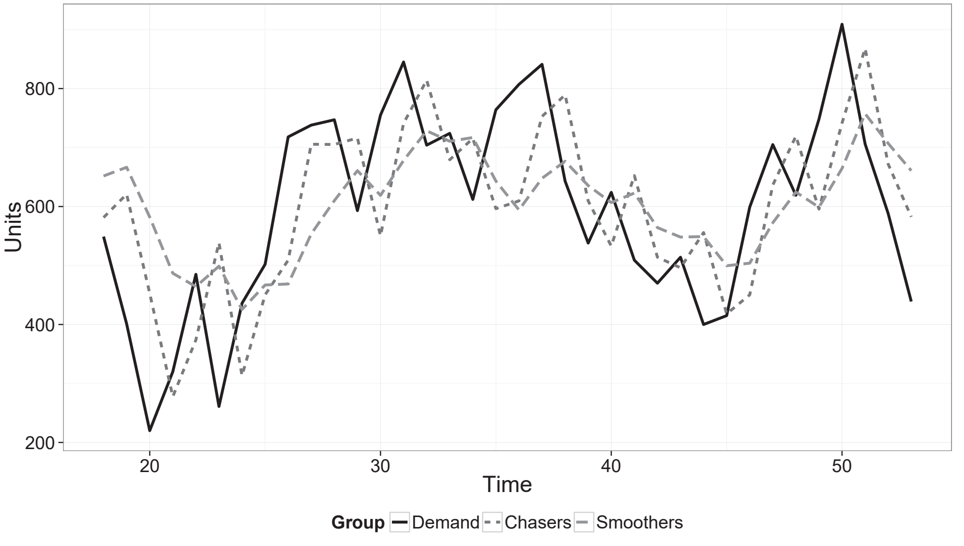
	Department incentive		Company incentive		
	Operations	Sales	Operations	Sales	Total
Group 1	63	69	68	71	271 (76%)
Group 2	22	20	24	20	86 (24%)
Total	85	89	92	91	357 (100%)

their forecast is inflated to the extent that it surpasses the just observed demand in time period 23. The same effect is observed in time periods 31, 34, and 36, albeit in opposite direction: the forecast is deflated to the extent that it is lower than the just observed demand. This shows that chasers are prone to seeing short-term trends. By contrast, Group 2 produces a substantially less volatile forecast series with gradual changes to the forecast, which are often of the same sign. Group 2 is only weakly influenced by forecast errors—these 'smoothers' overvalue their forecast, and strongly underreact to forecast errors. This behavior of smoothers is clearly distinct from the chasers' tendency to undervalue their forecast, overreact to forecast errors and heightened sensitivity towards short-term trends.

Figure 4.2 depicts the forecast series of chasers and smoothers after subtracting the optimal forecast in each time period $\hat{d}_{t|t-1}$, using (4.13), to derive the bias relative to the optimal forecast within the sample for each time period. It shows that for smoothers the first half of their consecutive forecasts has a positive bias, whereas the second half of their consecutive forecasts is characterized by a negative bias. In other words, by overvaluing their own forecasts and underreacting to forecast errors, smoothers thus generate forecasts that suffer from a consistent bias.

The different forecasting behavior of chasers and smoothers has substantial ramifications for forecast performance. Table 4.3 summarizes the performance measures for chasers, smoothers, and all participants combined. The RMSE for chasers, 170, is considerably and significantly smaller than that of smoothers, 216 ($p < 0.01$). Moreover, the relative bias of chasers and of all participants combined is close to zero, while smoothers have a substantial bias of 61% over average demand.

Figure 4.1: **Heterogeneity in judgmental forecasting**
This figure shows the average time series of forecasts of the two groups and the time series of the demand. Two different types of forecasting behavior can be seen: forecasters in Group 1 strongly adjust their forecasts, whereas forecasters in Group 2 weakly adjust their forecasts.



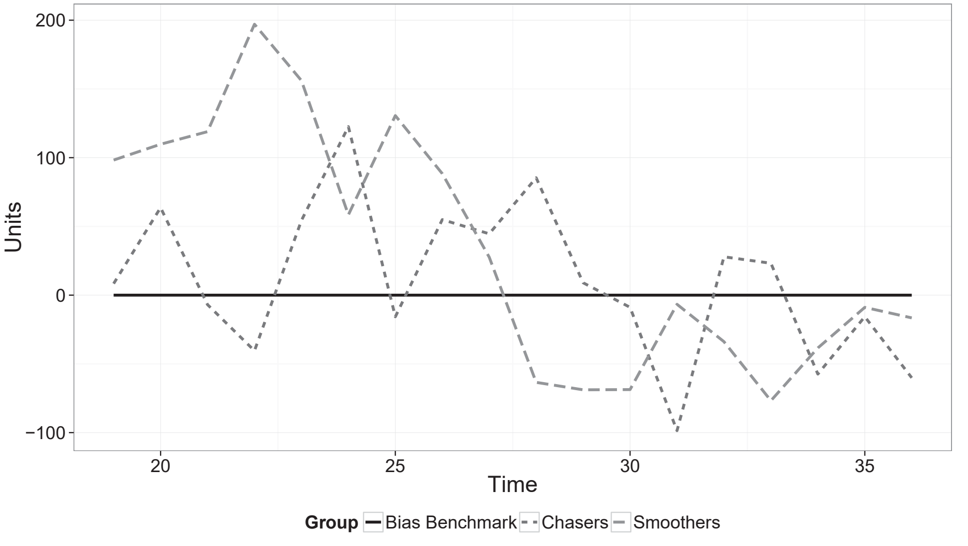
4.5.2 Generalized forecasting model

We estimate the generalized forecasting model with and without time-varying parameters for chasers and smoothers separately, and for all participants together to demonstrate the importance of heterogeneity for analyzing forecast behavior.

Table 4.4 gives the estimated results for the generalized forecasting model without time-varying parameters (4.8). The estimated smoothing parameter α of 0.70 for all participants is substantially and significantly higher than the optimal α of 0.61 ($p < 0.01$). The value is similar to the α of the descriptive measures mentioned by Kremer et al. (2011), and shows that forecasters overreact in a relatively unstable environment, thus seemingly contrasting the conclusion of Kremer et al. (2011) that forecasters underreact in relatively unstable environments.

Forecasting heterogeneity can explain the differences between these two conclusions. Observed forecasting heterogeneity plays a critical part in estimating the generalized forecasting model. The estimates of α in Table 4.4 for chasers and smoothers are substantially and significantly different, showing that the two groups respond dif-

Figure 4.2: **Different forecast bias for chasers and smoothers**
This figure shows the forecast series after subtracting the optimal forecast in each time series using (4.13) to derive the bias relative to the optimal forecast within the sample for each time period. Smoothers consistently forecast too high until period 29, after which they consistently forecast too low.



ferently to forecast errors. That is, chasers overreact to forecast errors as indicated by an estimated α of 0.78, whereas smoothers substantially underreact to forecast errors as expressed by an estimated α of 0.36. In other words, whereas most participants strongly overreact to forecast errors in this relatively unstable environment in a manner consistent with Kremer et al. (2011), our findings offer a further refinement of this established notion in the sense that a substantial portion of participants, the smoothers, intriguingly, underreacts.

Further evidence for distinct anchoring and adjustment behavior is contained in the estimated θ's, which generalize the anchor and adjustment model (4.4), by giving more flexibility to the form of the anchor. A θ equal to one means that the last forecast is used as anchor, whereas a value of zero means that a fixed long-term constant A is used as an anchor. The results in Table 4.4 again reveal distinct forecasting behavior for chasers and smoothers. The estimated θ of 0.93 (which is close to 1) indicates that chasers tend to anchor on the last forecast, whereas smoothers, with an estimated θ of 0.59, rather anchor on a mix of the last forecast and a fixed long-term constant. In effect, they smooth the anchor as well.

Table 4.3: **Forecast performance**
Forecast performance of participants with RMSE (4.10), Bias (4.11), and Relative Bias (4.12).

	RMSE	Bias	Relative bias
Chasers	169.794	-57.829	-0.100
Smoothers	216.071	354.986	0.616
All	179.808	23.115	0.040

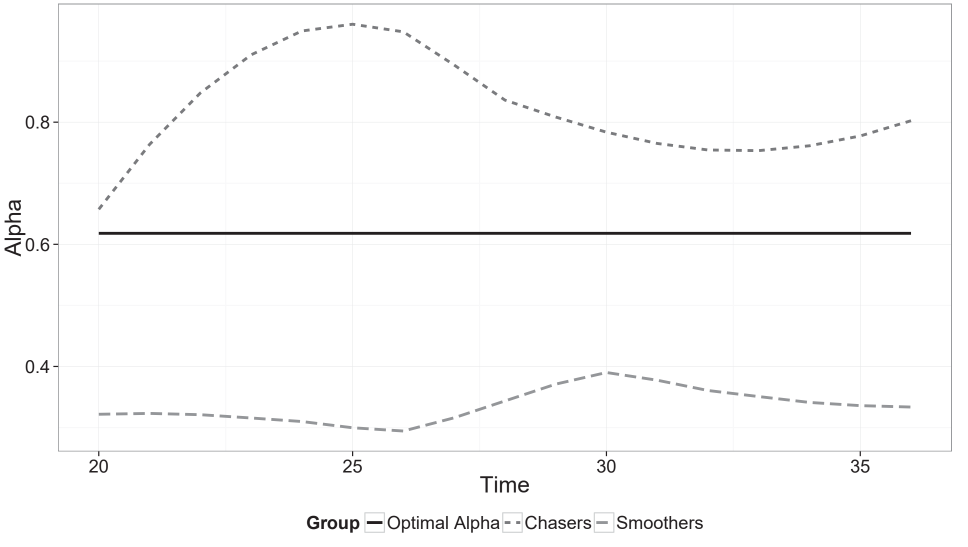
Table 4.4: **Estimates of the generalized forecasting model**
Estimates are given for different subsets of the data, either based on all participants combined, or on chasers and smoothers separately. Standard errors are in parentheses. The last column gives the differences between the estimates of the two groups (Δ(chasers, smoothers)), which are all significantly different from zero ($**$, $p < 0.01$).

	All (n = 357)	Chasers (n = 271)	Smoothers (n = 86)	Δ(chasers, smoothers)
α	0.698 (0.005)	0.780 (0.006)	0.365 (0.021)	0.416$**$
β	0.102 (0.009)	0.130 (0.010)	0.035 (0.039)	0.095$**$
θ	0.909 (0.002)	0.927 (0.001)	0.594 (0.128)	0.332$**$
A	577.792 (0.150)	573.161 (0.149)	596.869 (1.187)	-23.708$**$

The final parameter of interest of the generalized forecasting model is β, the influence of short-term trends in forecasts. For smoothers, the estimate is 0.03, which is not significantly different from 0, meaning that smoothers are generally not sensitive to seeing short-term forecast trends. For chasers, in contrast, the estimate of β is 0.13, which indicates that they are prone to seeing short-term trends in the data. As a result, chasers not only strongly overreact to forecast errors, but they even go as far as to overextend by acting on imagined trends where there is only noise.

A similar impact of heterogeneity is observed for learning effects of forecasting. Bostian et al. (2008) found that parameter values can change over time, thus signifying learning effects. Using their approach of linear change (4.6), we also find a small learning effect where participants slowly move towards the optimal with a 2% change in their α. By extending the model to include time-varying parameters (4.8) we can more flexibly trace how forecast behavior changes over time. In the extended model, parameters increasing or decreasing towards optimal values imply learning effects. Figure 4.3 shows how the conditional mean of smoothing parameter α changes over time for chasers and smoothers, respectively. The behavior of both groups is nonlinear, but the conduct of chasers changes much more dramatically over time than that of smoothers. Nonetheless, in contrast to the conclusions of Bostian et al. (2008), we find no evidence of learning effects for either of the two groups.

Figure 4.3: **Learning effects**
This figure shows the conditional means of α's for chasers and smoothers over time.



By way of a robustness check, we applied our approach to the experimental data consisting of 72 practitioners. Even though the analysis cannot be as extensive, because the number of participants is much lower, and, consequently, the standard error of results much higher, the results are, nonetheless, not materially different than the ones presented previously. Table 4.5 shows that the distinction between chasers and smoothers is equally apparent in the data set based on practitioners, although the smoothers represent 18% instead of 24% of the participants.

The evidence reported above bolsters our confidence in our estimated generalized forecasting model, which shows the existence of two distinct types of forecasters, based on groups of participants found using clustering. These two groups of chasers and smoothers, respectively, are substantially and significantly different in terms of behavior as captured by the generalized forecasting model. Not only do they differ in the extent to which they adjust their forecasts, but also in what they use as an anchor. Chasers strongly overreact to forecast errors, and tend to perceive short-term trends in the demand series. Smoothers strongly underreact to forecast errors, and

Table 4.5: **Estimates of the generalized forecasting model for 72 practitioners**
Estimates are given for different subsets of the data, either based on all participants, only on chasers, or only on smoothers. Standard errors are in parentheses.

	All (n = 72)	Chasers (n = 59)	Smoothers (n = 13)
α	0.644 (0.215)	0.776 (0.244)	0.244 (0.420)
β	0.088 (0.538)	0.146 (0.290)	0.154 (0.890)
θ	0.951 (0.091)	0.950 (0.052)	0.682 (0.562)
A	571.951 (5.169)	572.753 (4.927)	568.312 (18.486)

even smooth their anchor as a combination between their last forecast and a fixed long-term constant. We find no evidence for learning effects.

4.5.3 Departmental biases: roles and incentives

So far we have explored the existence and nature of distinct types of forecasting behavior. We noted that the observed behavioral differences are independent of departmental roles and incentives (see Table 4.2) and that smoothers generate forecasts that suffer from a consistent bias (see Figure 4.2), independent of their role and incentive. This led us to conclude that we should analyze the forecast behaviors for chasers and smoothers separately. Here, we turn to the consequences of departmental roles and incentives for forecast performance.

Table 4.6 summarizes the effects of departmental roles (operations vs. sales), incentives (department vs. company), and of the type of computer agent (neutral vs. other department) on participants' forecast behavior over each condition, by giving the estimated bias of the forecast. We examine chasers and smoothers separately, and for each group list the bias per role, incentive, and type of agent the participant is paired with. When paired with a neutral agent, chasers with department incentives display a negative bias of -63.1 if they are operations managers, or a positive bias of 27.5, if they are sales managers. Smoothers in the same conditions also display a negative bias of -69.7, if they are operations managers, or a positive bias of 88.3, if they are sales managers. Table 4.6 shows that these differences remain if we ignore the role specific incentive. The estimated biases substantially and significantly differ between the two roles, so that assigning roles has a strong impact. That is, participants with an operations role have a negative bias in their forecasts, while those in a sales role have a positive bias, even if their incentive is to minimize the forecast error.

Table 4.6: **Departmental biases**
Estimates of the forecast bias in the generalized forecasting model (4.8), seen as an adjustment inflating or deflating the forecast, per time period from the generalized forecasting model for the effects of roles, incentives and role of computer agent for chasers and smoothers. Standard errors are in parentheses. The columns first differentiate between department and company incentive, followed by the role of the participant. The rows first differentiate between chasers and smoothers, followed by the role of the computer agent. Strong effects are present for both chasers and smoothers. Participants generate biased forecasts because of their roles, even without incentives.

	Department incentive		Company incentive	
	Operations	Sales	Operations	Sales
Chasers				
Neutral agent	-63.068 (6.961)	27.526 (3.587)	-71.634 (7.679)	29.781 (4.387)
Agent other dep.	57.000 (4.188)	-103.788 (9.146)	31.600 (3.214)	-142.586 (5.442)
Smoothers				
Neutral agent	-69.732 (19.241)	88.337 (23.006)	-52.831 (14.051)	41.780 (16.595)
Agent other dep.	77.516 (21.055)	-90.219 (27.864)	35.883 (7.935)	-93.089 (17.003)

Table 4.6 further shows an interesting change of signs when participants are paired with an agent from the other department. The biased computer agent in that case does not share a mean forecast, but rather provides an adjusted forecast based on its own role, which may cause participants to display a stronger bias. More specifically, chasers with an operations role have a bias equal to either −63.1 or 57.0 depending on whether they are paired with a neutral agent or a computer agent from the other department. A similar effect is observed over all conditions. The bias of the computer agent thus has the effect of increasing or decreasing forecasts to the extent that it switches the sign of the bias.

4.6 Discussion and conclusion

Our study has demonstrated that forecasting heterogeneity matters. Forecasting behavior differs systematically between individuals to the extent that two markedly different types of forecasters can be distinguished. One is characterized by overreaction to forecast errors and has been labeled chasers, while the other is characterized by underreaction to forecast errors, and has been labeled smoothers. The existence of two distinct groups is highly relevant for the analysis. Results obtained from earlier research on individual biases is possibly misleading as they are based on aggregate results and ignore systematic behavioral differences. This explains why our finding that forecasters overreact in a relatively unstable environment conflicts with the conclusion of Kremer et al. (2011). The difference between chasers and smoothers, and nonlinear behavior, also explains why we find no evidence for learning effects in contrast to Bostian et al. (2008).

Extending the models used in earlier behavioral experiments, we propose an approach relying on wavelets and state space modeling to capture individual forecasting behavior. Our empirical estimations of the state space model of the two groups found using wavelets and clustering, show that the two types of chasers and smoothers exhibit substantially and significantly different behavior. Chasers not only strongly overreact to forecast errors, but are also prone to seeing short-term trends. Smoothers not only underreact, but are fundamentally different under the anchor and adjustment model, as they use a smoothed value as anchor instead of the last forecast.

Furthermore, we demonstrate the existence of persistent departmental biases of roles and incentives. In line with conclusions of Kuo and Liang (2004) and Önkal et al. (2012), we find that forecasting behavior is influenced by roles. In contrast to the conclusion of Yaniv (2011), the effect of roles is not negated using incentives. We are unable to differentiate between intentional and unintentional biases, as roles have a strong effect, even without incentives, which has ramifications as we can no longer assume that we can disentangle the two (e.g. Oliva and Watson, 2009, 2011).

Our findings are also important for practice, as forecast behavior directly affects forecast performance, which can have large financial ramifications. Chasers and smoothers have substantially different forecast performance, so that recognizing the difference between these two types of forecasting can lead to better hiring and training practices for forecasters. The impact of departmental biases also has ramifications for how the forecasting process is orchestrated within companies when multiple departments participate, as roles and the behavior of other participants affect behavior.

Different types of forecasting behavior will remain an important topic for future study, as they impact both research done so far and practice. The novel methodology we outlined here, relying on wavelets and state space modeling, should prove to be flexible in similar types of research.

Chapter 5

Coordinating Judgmental Forecasting: Coping with Intentional Biases

Co-authors: J. van Dalen and L. Rook

Abstract

Biases in judgmental forecasting have often been studied, but unintentional and intentional biases have never been disentangled. We isolate intentional biases in the context of departmental roles and incentives in corporate forecasting processes. Through an experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, we examine the effects of roles, incentives, and various weighing schemes on behavior and performance. We find that roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We test the claim that an accuracy-weighted scheme can remove unintentional biases, and conclude that though this halves these biases, it does not fully remove them. Finally, we observe that a weighing scheme that explicitly corrects biased inputs shows great promise in reducing intentional as well as unintentional biases. In our experiment, this scheme reduces biases by 35%. Our work shows the importance of disentangling intentional and unintentional biases for research, and our insights have substantial ramifications for the design of the forecasting process in terms of coordination mechanisms and incentives by quantifying the impact of roles and incentives.

Keywords: judgmental, biases, decision-making, incentives, negotiation.

5.1 Introduction

Judgmental forecasting is commonly used in practice and affects company performance (Fildes et al., 2008; Syntetos et al., 2011, 2010). Inaccurate demand forecasts can have substantial financial ramifications. Involving multiple organizational departments, such as sales, operations, and finance, in generating forecasts, often embedded in sales and operations planning (S&OP) (Singhal and Singhal, 2007), has been reported to reduce inaccuracies (Oliva and Watson, 2009, 2011). However, when departments have different financial risks, they do not necessarily share the same goal of minimizing forecast errors, resulting in suboptimal financial performance for the company (Nauta and Sanders, 2001).

Involving multiple departments affects forecasting mainly through roles and incentives (Oliva and Watson, 2009, 2011). Roles contextualize tasks and may lead to biases: inflated or deflated forecasts (Kuo and Liang, 2004; Önkal et al., 2012). Nauta and Sanders (2001) and Nauta et al. (2002) mention operations and sales as an example, arguing that organizational departments commonly have opposing interests: while the operations department focuses on efficiency and costs, the sales department prioritizes customer service and sales development. Incentives steer forecasting behavior to the extent that managers deliberately increase forecast biases by adjusting forecasts (Oliva and Watson, 2009, 2011). Because roles and incentives provide context, people are both unknowingly influenced by them and act deliberately upon them. Inflated and deflated forecasts arise because of both unintentional as well as intentional forecasting behavior.

The interplay of unintentional and intentional forecasting behavior has been largely ignored in the literature (Oliva and Watson, 2009, 2011), which is not surprising, considering its complexity. Judgment is rife with inherent biases of human decision-making, leading to unintentional forecast biases (Lawrence et al., 2006). Groups, such as organizational departments, are exposed to the same routines and systematic decision-making errors as individuals (Kerr and Tindale, 2011). In addition to unintentional biases, intentional adjustments affect forecasts, leading to intentional forecast biases. Because in practice intentional adjustments are not observed in isolation from unintentional adjustments, we are unable to disentangle intentional biases from unintentional biases. Consequently, we do not fully understand the effects of roles and incentives on behavior and performance.

Oliva and Watson (2009, 2011) illustrate the importance of roles and incentives for the forecasting process by describing the overhaul of the forecast process at Leitax, a manufacturer of consumer electronics. Prior to the change, the forecasting process was fragmented over departments. Sales shared their generated forecasts informally with operations and finance. Operations required forecasts for purchasing and production decisions; finance required them for financial planning and management. Not convinced of the adequacy of each other's forecasts, they generated their own demand forecasts, resulting in large financial losses. Centralizing and redesigning the process was successful and had a major impact on the operations of Leitax. The forecast accuracy increased dramatically by 30 percentage points from 58% to 88%, which entailed millions of savings in inventory.

Though the Leitax case demonstrates the importance of design choices for the forecasting process, it offers no insight into how the design affects intentional and unintentional behavior. In the redesigned forecast process, the separate forecasts of sales, product planning and strategy, and demand management together determine the final consensus forecast based on their past performance (Oliva and Watson, 2009, 2011). Other organizations also rely on this weighing scheme, such as Norges Bank and manufacturers of fast-moving consumer goods (Bjørnland et al., 2012; Protzner, 2015). Supposedly, this scheme removes the influence of roles and incentives, improving the forecast by negating intentional biases (Oliva and Watson, 2009, 2011). However, the effects of roles, incentives, and weighing schemes, on the actual behavior of managers in the forecasting process have not been examined.

In this chapter, we study intentional biases in the context of different departmental roles and incentives in corporate forecasting processes. Through an experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, we disentangle intentional from unintentional biases and examine the effects of roles, incentives, and various weighing schemes on behavior and performance. We find that roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We test the claim that an accuracy-weighted scheme can remove unintentional biases, and conclude that though this halves these biases, it does not fully remove them. Finally, we observe that a weighing scheme that explicitly corrects biased inputs shows great promise in reducing intentional as well as unintentional biases. In our experiment, this scheme reduces biases by 35%. Our work shows the importance of disentangling the two sources of biases for research, and our insights have substantial ramifications for the design of the forecasting process in terms of coordination mechanisms and incentives.

The remainder of this chapter is organized as follows. Section 5.2 outlines the relevant literature on intentional biases in judgmental forecasting and on weighing schemes for formulating consensus forecasts, and states our hypotheses. Section 5.3 specifies our experiment and our methods to examine participants' behavior. Section

5.4 lists the results and their implications, while Section 5.5 concludes and gives suggestions for future research.

5.2 Theoretical background

Intentional forecast biases are often overlooked in the literature, and never isolated from unintentional biases for study. Weighing schemes to combine forecasts are widely studied as mechanisms to improve forecast accuracy. Though their use supposedly removes intentional biases, their influence on forecasting behavior remains ignored. In this section, we formulate hypotheses and determine the objectives of our experiments to examine how the design of the forecasting process affects behavior and performance.

5.2.1 Intentional forecast biases

Forecasts can be subject to managerial pressure (Syntetos et al., 2009b), and are not necessarily supposed to minimize forecast biases. Organizations can maintain objectives other than forecast accuracy: forecast biases can be intentional (Lawrence et al., 2000) and a "result of deliberate and rational decision making behavior on the part of the forecasters" (Lawrence and O'Connor, 2005, p.3). Documented examples of intentional biases include forecasters who inflate forecasts to ensure that suppliers give them priority (Syntetos et al., 2009b) or to increase the publicity of the forecast (Ashiya, 2009). Wright and Rowe (2011, p. 12) conclude that "[w]e must always remember that forecasts are rarely, in themselves, disinterested and innocent products of the group process in which they are produced and this reality should cause us to reconsider the way in which we evaluate forecasts."

Departmental roles and incentives are sources of unintentional and intentional forecast biases. Yaniv (2011) fully ascribes biases to incentives, ignoring roles, and concludes that forecasting behavior differs substantially between departments only when financial incentives differ. By contrast, Kuo and Liang (2004) highlight the importance of roles. They conclude that departmental roles affect forecasting behavior, even when forecasters receive exactly the same information and have no role-specific incentives or interests, illustrating the unintentional bias provided by roles. Likewise, Önkal et al. (2012) find significant effects on the forecast of varying roles, even without incentives. Participants with the role of forecasting executive, marketing director, or production director display a strong commitment to their own roles (Önkal et al., 2012). Yaniv (2011), Kuo and Liang (2004), and Önkal et al. (2012) do not differentiate between intentional and unintentional biases. However, our previous study (see Chapter 4) exploring the interaction of roles and incentives, demonstrates that both roles and incentives have a substantial influence on behavior and cause unintentional biases.

Forecast processes at organizations involve multiple stakeholders, which does not necessarily contribute to the quality of the forecasts, and possibly impairs forecasts due to the negative effects of how the group is organized (Brockhoff, 1983). Forecast accuracy depends on the group size, the way members interact, the performance of individual members, and shared representations of the forecasting task (Graefe and Armstrong, 2011; Kerr and Tindale, 2011). Alternatively, instead of groups that generate a single forecast, the separate individual forecasts of group members can be combined into a final forecast by using a weighing scheme.

5.2.2 Weighing schemes for combining forecasts

Weighing schemes for combining forecasts from forecasting methods, rather than individual people, have been extensively studied. A combination of forecasts obtained from various forecasting methods can reduce the forecast error, by being more robust to particular assumptions and wrong inferences (Bates and Granger, 1969; Diebold and Pauly, 1987). Indeed, empirical results show that a combination of separate forecasts often substantially improves forecast accuracy (Chan et al., 1999; Clemen, 1989; Diebold and Pauly, 1987). A simple average of forecasts can outperform separate forecasts (Fang, 2003), and is generally more robust than a weighted average (Palm and Zellner, 1992). Adding additional forecasts as inputs further improves accuracy (Makridakis and Winkler, 1983). Recent research continues to show the benefits of combining forecasts (Costantini and Kunst, 2011; Kurz-Kim, 2008; Rapach and Strauss, 2008; Swanson and Zeng, 2001; Wichard, 2011).

These results appear to hold when inputs are provided by judgmental forecasters instead of forecasting methods, in a context without roles and incentives (Ashton and Ashton, 1985; Clemen and Winkler, 1999; Lipscomb et al., 1998; Morris, 1977; Önkal et al., 2011). Whether the results change when forecasters do have varying roles and incentives is unknown. Though the claim that weighing schemes can negate the intentional biases of roles and incentives is appealing (Oliva and Watson, 2009, 2011), there is no evidence to support this.

Various weighing schemes to combine separate forecasts, in addition to the simple average, have been proposed. The accuracy-weighted scheme, a popular method, derives from the variance-covariance method. This method incorporates the accuracy of the individual forecasts, reflected by the variance of individual forecast errors, as well as the dependence between forecasts, reflected by the covariance between individual forecast errors (Winkler and Makridakis, 1983). Weights are calculated by means of linear regression (Granger and Ramanathan, 1984), principal component regression (Chan et al., 1999), or Bayesian shrinkage (Anandalingam and Chen, 1989; Diebold and Pauly, 1990; Min and Zellner, 1993; Walz and Walz, 1989). The accuracy-weighted scheme, seen at Leitax (Oliva and Watson, 2009, 2011), Norges Bank (Bjørnland et al., 2012), and manufacturers (Protzner, 2015) ignores the covariance between forecast errors to increase forecast accuracy, because of the sensitivity of the covariance to the sample cross-correlations, which results in highly unstable estimates of the weights (Clemen and Winkler, 1986; Makridakis and Winkler, 1983; Newbold and Granger, 1974; Winkler and Clemen, 1992).

5.2.3 Hypotheses

Roles and incentives affect behavior, and engender both unintentional and intentional forecast biases. We posit that these biases can be separated by distinguishing between a private forecast and a shared production quantity, determined sequentially by forecasters. The private forecast contains unintentional biases. The intentional bias is measured as the difference between proposed production quantities and private forecasts, disentangled from unintentional forecasting biases.

Previous research ascribes intentional biases solely to financial incentives (Oliva and Watson, 2009, 2011). But roles, even when unconnected to rewards or penalties, can affect intentional biases, by implying goals. Because prior research shows that roles and incentives both cause unintentional biases (see Chapter 4), we hypothesize that both also cause intentional biases and substantially impact performance.

Hypothesis 1. *Organizational roles, even without role-specific financial incentives, entail intentional biases.*

Hypothesis 2. *Financial role-specific incentives result enlarge intentional biases induced by organizational roles.*

To analyze forecasting behavior under varying conditions, we allow for heterogeneity among forecasters, as previous work shows that distinct types of forecasting behavior exist, labelled chasers and smoothers (see Chapter 4). Nevertheless, as the unintentional biases caused by roles and incentives are independent of the type of behavior (see Chapter 4), we hypothesize that intentional biases are similarly independent of the type of forecasting behavior.

Hypothesis 3. *Intentional biases are not related to the distinct types of forecasting behavior labelled chasers and smoothers.*

Weighing schemes supposedly negate the effects of roles and incentives, improving forecasts by removing intentional biases (Oliva and Watson, 2009, 2011). However, no studies examine the forecasting behavior under particular weighing schemes. Here, we test whether the accuracy-weighted scheme reduces intentional biases relative to using the simple average.

Hypothesis 4. *The accuracy-weighted scheme to combine separate forecasts results in lower intentional biases than the simple average.*

The different combination schemes discussed do not incorporate interaction between forecasters for a single decision. Yet, forecast performance hinges on how members interact (Graefe and Armstrong, 2011; Kerr and Tindale, 2011). Nauta and Sanders (2001) and Nauta et al. (2002) mention operations and sales as an example of organizational departments that commonly have opposing interests. By allowing revisions in response to other inputs and by having roles with opposing interests, a weighing scheme resembles a negotiation, which is likely to emphasize the competitive nature of the process and increase intentional biases.

Hypothesis 5. *Incorporating interactions between members in the forecast meeting increases intentional biases.*

If unintentional biases are reduced, but not removed, by the accuracy-weighted scheme, the extended weighing scheme of Palm and Zellner (1992), which explicitly models and corrects biases of inputs, potentially increases performance, especially because unintentional biases due to roles and incentives are present (see Chapter 4). Performance can also deteriorate because of misspecification of the scheme and estimation errors. Because the scheme is not as simple to interpret, its effect on behavior is not examined. However, combining inputs post-hoc using the scheme may demonstrate its potential value.

Hypothesis 6. *A weighing scheme that corrects inputs for biases outperforms weighing schemes without such a correction.*

5.3 Experimental design and data

We conducted an experiment to examine our hypotheses about forecaster behavior under various roles, incentives, and weighing schemes. After studying a time series of 18 periods of historic demand, participants separately provide a forecast and production quantity for the next period. They subsequently see the actual outcome and updates of the available information—profit, lost sales, obsolescence, and forecast accuracy. They repeat these tasks and see the outcome successively for 18 periods.

The forecasts are neutral, representing the participants' expected demand, and are separate from the desired production quantities. The task is sequential: participants privately forecast the demand for the next period, after which they propose a production quantity. Another manager, represented by a computer agent, simultaneously proposes a production quantity. Both of these quantities are used in various weighing schemes to determine the production decision. The agent allows us to simulate interdepartmental decision-making in a fully controlled environment. In this setup, the intentional bias is isolated for analysis by defining it as the difference between the demand forecast and the proposed production quantity.

The experiment has a two-by-two-by-two-by-two mixed factorial design, varying with respect to the departmental role (sales or operations), incentive scheme (absence or presence of a role-specific financial incentive), type of computer agent (absence or presence of a role-specific financial incentive), and weighing scheme used (accuracy-weighted or interaction). The weighing scheme that defines the final production quantity varies between distinct phases of the experiment. Apart from the initial training phase, the experiment consist of two phases. The first phase uses a simple average of inputs, whereas the second phase uses either the accuracy-weighted scheme or allows for interaction between participant and agent to determine the final production quantity.

Below, we specify the experiment by detailing the forecast decision, the production quantity decision, the roles and incentives of participants and agents, and the weighing schemes, after which we introduce our sample and measures of analysis. The decision-making context follows from previous behavioral experiments: the forecasting decision is derived from the study of Kremer et al. (2011) and the production quantity decision from Schweitzer and Cachon (2000).

5.3.1 The forecast decision

Similar to Kremer et al. (2011), we use a local level model, also known as a random walk with noise, to simulate the demand:

$$\begin{aligned} d_t &= l_t + \varepsilon_t, \qquad & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ l_t &= l_{t-1} + \nu_t, \qquad & \nu_t &\sim N(0, \sigma_\nu^2) \end{aligned} \tag{5.1}$$

In our case, σ_ν^2 and σ_ε^2 are set equal to 100, thus implying equal parts of signal and noise.

The optimal forecast

In the study of Kremer et al. (2011), the optimal forecast is used to compare performance between participants. However, we only use it to specify agent behavior. For a local level model, single exponential smoothing, which updates the forecast based on the observed forecast error, is the optimal forecast method (Lawrence and O'Connor, 1992):

$$\hat{d}_{t+1|t} = \hat{d}_{t|t-1} + \alpha(d_t - \hat{d}_{t|t-1}) \tag{5.2}$$

where $\hat{d}_{t+1|t}$ denotes the forecast of demand for period $t+1$ at time t, and d_t denotes the observed demand at time t; α is a smoothing parameter.

The smoothing parameter α^* that minimizes the mean squared forecast error depends on the signal-to-noise ratio $q = \sigma_\nu^2/\sigma_\varepsilon^2$. Applying the Kalman filter to the local level model (5.1) leads to equation (5.2) and gives the optimal smoothing parameter from the steady state of the Kalman gain, the optimal weighing factor for new information (Durbin and Koopman, 2012), as:

$$\alpha^* = \frac{\sqrt{q(q+4)} - q}{2} \tag{5.3}$$

Agents use single exponential smoothing with (5.3) to forecast demand for the next period.

5.3.2 The production quantity decision

The production quantity decision is based on the newsvendor model, following the behavioral experiment of Schweitzer and Cachon (2000), in which a production quantity $q_{t+1|t}$ needs to be decided at time t for sale during the next period $t+1$. The produced quantity is only available during time period $t+1$. After observing the outcome in period $t+1$, in which lost sales are directly observed, a new production quantity has to be set for period $t+2$.

Profit is determined by the revenue p for each unit sold minus the cost c for each unit produced. The number of units sold is equal to the minimum of the produced quantity and demand. Given production quantity q and demand d, profit $\pi(q, d)$ is:

$$\pi(q, d) = p \min(q, d) - c \cdot q \tag{5.4}$$

Expected profit for the demand distribution F, with density f, is:

$$E[\pi(q,d)] = [1 - F(q)]\pi(q,q) + \int_0^q f(x)\pi(q,x)dx \tag{5.5}$$

The optimal production quantity

In the study of Schweitzer and Cachon (2000), the optimal production quantity is used to normatively assess the performance of participants. However, similar to the optimal forecast, we only use it to specify agent behavior. The optimal production quantity, q^*, maximizes expected profit (5.5) by balancing the costs of lost sales $(p - c)$ and the total cost (p) of being either overstocked (c) or understocked $(p - c)$, and follows from:

$$F(q^*) = \frac{p-c}{c+(p-c)} = \frac{p-c}{p} \tag{5.6}$$

which is referred to as the critical fractile (Schweitzer and Cachon, 2000).

The optimal order quantity is based on this critical fractile rather than on the expected demand. Applying the Kalman filter to the local level model (5.1) gives an expression for the variance of the demand. Since the forecast distribution is normal (Durbin and Koopman, 2012), the mean and the variance characterize the entire distribution. The expression for the prediction error variance derived from the Kalman filter consists of the variance of the next state plus the variance of the noise:

$$\mathrm{Var}(\hat{d}_{t+1|t}) = \mathrm{Var}(\hat{l}_{t+1|t}) + \sigma_\varepsilon^2 \tag{5.7}$$

Similar to calculating the optimal smoothing value, the variance of the state, $\mathrm{Var}(\hat{l})$, has a steady state, satisfying the following equation, derived from applying the recursive Kalman filter:

$$\begin{aligned}\mathrm{Var}(\hat{l}) &= \mathrm{Var}(\hat{l})\left(1 - \frac{\mathrm{Var}(\hat{l})}{\mathrm{Var}(\hat{l}) + \sigma_\varepsilon^2}\right) + \sigma_\nu^2 \\ &= \sigma_\varepsilon^2(q + \sqrt{q^2 + 4q})/2\end{aligned} \tag{5.8}$$

So, the prediction error variance is:

$$\mathrm{Var}(\hat{d}_{t+1|t}) = \sigma_\varepsilon^2(q + \sqrt{q^2 + 4q})/2 + \sigma_\varepsilon^2 \tag{5.9}$$

The expressions for the mean forecast (5.2) and the variance (5.9) together characterize the normal demand distribution F in (5.5) and (5.6):

$$d_{t+1|t} \sim N(\hat{d}_{t+1|t}, \mathrm{Var}(\hat{d}_{t+1|t})) \tag{5.10}$$

We can now solve for q^* using the inverse distribution function and the critical fractile:

$$q^* = F^{-1}\left(\frac{p-c}{p}\right) = \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})} \cdot \Phi^{-1}\left(\frac{p-c}{p}\right) \tag{5.11}$$

which gives the optimal order quantity, based on the optimal forecast and the trade-off of being either over- or understocked. Agents derive the optimal order quantity (5.11) using the optimal forecast and the steady state values of (5.9).

5.3.3 Roles and incentives

To examine behavior under different roles and incentives, participants have the role of either operations or sales managers. The operations department focuses on production and inventory levels; the sales department focuses on product availability. Incentives penalize participants for outcomes straying from their department's objective, minimizing either obsolescence or lost sales, or from the company's objective, maximizing profit and minimizing ex-post inventory error.

Under incentives for the company's objective, which are not role-specific, the sales price p is 2 Euro and the cost of production c is 1 Euro. The optimal order quantity $q^*_{t+1|t}$ at time t, following (5.11), is equal to the expected demand $\hat{d}_{t+1|t}$, regardless of the departmental role, because of a symmetrical cost structure of lost sales and obsolescence. In this case, there is no incentive for an intentional bias.

However, under incentives for the department's objective, the cost structure is asymmetrical. The sales price p remains 2 Euro. Operations managers are penalized for obsolescence, doubling the associated cost; sales managers are similarly penalized for lost sales. This shifts the trade-off of the costs of lost sales and the total cost of being either overstocked or understocked, changing the critical fractile (5.6) for operations managers and sales managers respectively to:

$$\begin{aligned} F(q^*_{\text{operations}}) &= \frac{p-c}{2c+(p-c)} \\ F(q^*_{\text{sales}}) &= \frac{2(p-c)}{c+2(p-c)} \end{aligned} \tag{5.12}$$

As a result, for $p = 2$ and $c = 1$ the optimal order quantities q^* (5.11) for operations managers and sales managers respectively become:

$$\begin{aligned} q^*_{\text{operations}} &= \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})}\Phi^{-1}\left(\frac{1}{3}\right) \\ q^*_{\text{sales}} &= \hat{d}_{t+1|t} + \sqrt{\mathrm{Var}(\hat{d}_{t+1|t})}\Phi^{-1}\left(\frac{2}{3}\right) \end{aligned} \tag{5.13}$$

The incentives of these two roles are symmetrical around the expected demand. If a simple average is taken of the two optimal order quantities and if behavior is rational and based on unbiased forecasts, the effects of the incentives cancel out.

The change in the optimal order quantity illustrates the effect of the role-specific department incentive. In addition, (5.11) and (5.13) specify the desired production quantity of agents with a company objective (not role-specific) or departmental objective (role-specific), respectively.

To simulate interdepartmental decision-making between sales and operations, the production decision in each time period is based on the shared quantity inputs of the participant and the computer agent. The role of the agent is complementary to the participant's role, i.e. the computer agent takes on the role of an operations manager when a participant has the role of a sales manager. The agent provides its desired production quantity as input.

5.3.4 Weighing schemes

Three different weighing schemes are used in different phases of the experiment to combine inputs from the participant and the agent into a final production quantity. The weighing scheme used in the first phase of the experiment is a simple average of inputs, calculating the production decision q_o, based on the proposed quantities of participant r and agent a, in each time period t, as:

$$q_{o,t} = (q_{r,t} + q_{a,t})/2 \tag{5.14}$$

The second combination scheme is the accuracy-weighted scheme. It is defined as a weighted average based on the past performance of the participant and agent, in which the covariance is ignored, as is commonly done in practice (Bjørnland et al., 2012; Clemen and Winkler, 1986; Winkler and Clemen, 1992):

$$\begin{aligned} q_{o,t} &= w_{r,t} q_{r,t} & &+ \quad w_{a,t} q_{a,t} \\ w_{r,t} &= \frac{e_{a,t}}{e_{r,t} + e_{a,t}} & w_{a,t} &= \frac{e_{r,t}}{e_{r,t} + e_{a,t}} \\ e_{r,t} &= \frac{1}{t-1} \sum_{i=1}^{t-1} (q_{r,i} - d_i)^2 & e_{a,t} &= \frac{1}{t-1} \sum_{i=1}^{t-1} (q_{a,i} - d_i)^2 \end{aligned} \tag{5.15}$$

Note that $q_{o,t}$ is a convex combination of the inputs based on the observed forecast accuracy up to time period t.

The third combination scheme has interaction in each time period by allowing the participant and the agent to revise their inputs after seeing the other's input.

By having roles with opposing interests, the weighing scheme resembles a negotiation. Interaction is limited to four rounds. This should be sufficient to determine whether the scheme affects participant behavior. In each of these rounds, the agent and the participant simultaneously propose production quantities. After seeing each other's proposals, they can update their quantity for the next round. There are no restrictions for the participants: they can increase, decrease, or leave the proposed quantity unchanged. The agent's behavior is outlined below. After the last round of interaction, the average of the last two inputs of the agent and the participant determines the production outcome.

The agent follows a simple algorithm during interactions, which includes random variation to avoid deterministic behavior. The agent neither behaves competitively nor punishes the participant: it can either adjust its input towards the participant's or leave it unchanged. The agent's desired production quantity is its input for the first round of the negotiation. Also, the agent never moves outside of the range set by the 5th and 95th percentiles of the forecast distribution, to limit its reaction to possibly extreme inputs by participants.

In the second round of the interaction, the agent adjusts its quantity, reducing the gap between its own and the participant's quantity. Its behavior in the third and fourth round depends on the preceding actions of participants. If the participant's quantity is not closer towards the agent's proposal, the agent does not adjust its proposal in return. However, if the participant decreases the gap, the agent further adjusts its proposal towards the participant's.

If the agent adjusts its quantity, it adjusts it by one third of the distance between its own most recent proposal q_a and the latest proposal of the participant q_r, representing a substantial step towards the participant's quantity. It then increases or decreases, depending on the participant's role, its proposal by this quantity times a random factor, to include variation in its new proposal q_a and avoid deterministic behavior, which can be quickly learned by participants:

$$\begin{aligned} x &\sim \text{Beta}(2,2) \\ q_a &:= q_a + \frac{q_r - q_a}{3}(0.5 + x) \end{aligned} \tag{5.16}$$

This proposed beta distribution is attractive, because it restricts x to $0 \leq x \leq 1$, the expected mean and mode is 0.5, and the probability mass is highest at the mean after which it tapers off for higher or lower values. As skewness is zero, the distribution is symmetrical around the mean.

5.3.5 Samples

We generated two time series of demand, one for each phase, using (5.1) for 36 time periods. The first half of each time series serves as the historic data for participants; the second half is used for the decision-making.

For the condition of the accuracy-weighted weighing scheme, analyses are based on 357 participants (240 men and 117 women with an average age of 21) who are randomly allocated over the four different conditions of role (sales or operations) and incentive (not role-specific company incentive or role-specific department incentive). The number of participants for each role and incentive is listed in Table 5.1. Students of a Business Administration program participated in 2013 as part of their course. They were familiar with the topic. Behavioral experiments are commonly conducted with students to ensure that analyses are based on a large number of participants (e.g. Bolton and Katok, 2008; Bostian et al., 2008; Kremer et al., 2011; Schweitzer and Cachon, 2000). Typically, experienced managers and students exhibit the same behavior (Bolton et al., 2012).

Table 5.1: **Experimental data over the four conditions for the two conditions of the weighing scheme**
The four experimental conditions are based on the two roles of operations and sales, and the incentives of either department or company.

	Company incentive (Not role-specific)		Department incentive (Role-specific)	
	Operations	Sales	Operations	Sales
Accuracy-weighted				
Total: 357	92	91	85	89
	(26%)	(25%)	(24%)	(25%)
Interaction				
Total: 72	16	19	19	18
	(22.22%)	(26.39%)	(26.39%)	(25.00%)

For the condition using interaction as a weighing scheme, the analysis is based on practitioners instead of students. The 72 practitioners, of which 51 men and 21 women, with an average age of 34, from various manufacturing companies, are all involved in forecasting or demand planning. Table 5.1 lists the number of participants for each role and incentive.

We exclude respondents who do not correctly answer the control questions, which check whether participants remember their role and incentive scheme at the end of the experiment, and understand the forecasting task of the experiment. We also leave out respondents who make typographical errors during the experiment. This is necessary because a simple input error influences subsequent rounds as the decisions are dependent on each other through time. Hence, we cannot simply correct obvious errors, interpolate, or treat particular inputs as missing values. No other selection criteria are applied.

5.3.6 Analyses

The main measure of interest is the intentional bias, defined as the difference between participants' forecast $d_{i,t}$ and their production quantity $q_{i,t}$:

$$\delta_{i,t} = \begin{cases} d_{i,t} - q_{i,t} & \text{if } i \text{ has an operations role} \\ q_{i,t} - d_{i,t} & \text{if } i \text{ has a sales role} \end{cases} \tag{5.17}$$

We define the intentional bias separately for operations and sales roles to ensure that the intentional bias follows from the context: a positive bias for operations means that the forecast is deflated, whereas a positive bias for sales means that the forecast is inflated. For interactions, the final proposed quantity is used.

Preliminary insight into the effect of different incentives and roles is given by graphing the average of δ_t per incentive type over time. In addition, the intentional bias $\delta_{i,t}$ is modeled as an AR(1) process with random slopes α_i and coefficients ϕ_i to incorporate heterogeneity using maximum likelihood. A dummy variable v indicates the absence (0) or presence (1) of the role-specific department incentive for participants.

$$\begin{aligned} \delta_{i,t} &= \alpha_i + \phi_i \delta_{i,t-1} + \beta v_i + \eta_{it} \\ \eta_{it} &\sim N(0, \sigma_\eta^2) \\ \alpha_i &\sim N(\mu_\alpha, \sigma_\alpha^2) \\ \phi_i &\sim N(\mu_\phi, \sigma_\phi^2) \end{aligned} \tag{5.18}$$

If $|\phi_i| < 1$, the autoregressive process is stationary, and the mean of δ_i is:

$$E[\delta_i] = \frac{\alpha_i + \beta v_i}{1 - \phi_i} \tag{5.19}$$

A positive mean for participants with the company incentive ($v_i = 0$) gives evidence for hypothesis (1) that roles, even without role-specific financial incentives, entail intentional biases. A β that is substantially and significantly higher than 0 gives evidence for hypothesis (2) that role-specific incentives result in larger intentional biases.

Model (5.18) is estimated separately for the two distinct types of forecasting behavior, chasers and smoothers. Estimates that do not differ significantly support hypothesis (3), which posits that intentional biases are not related to the distinct types of forecasting behavior labelled chasers and smoothers.

Behavior under the different weighing schemes is compared using the mean average intentional bias. In the first phase, the simple average is used. In the second phase, either the accuracy-weighted scheme or interactions between the participant and agent, allowing them to revise their inputs, is used. The change in the intentional bias in the second phase is tested using a Wilcoxon signed-rank test, to determine whether intentional biases are lower under the accuracy-weighted scheme, as posited by hypothesis (4), and whether intentional biases are higher when there is interaction between participant and agent, as posited by hypothesis (5).

If the intentional biases are not fully removed by the accuracy-weighted scheme, a possible alternative is offered by a scheme which de-biases the inputs. If the biases are constant and do not cancel out, the inputs can be explicitly de-biased by estimating constant biases θ for each input. Given θ_r and θ_a, the weights can be calculated as (Palm and Zellner, 1992):

$$\begin{aligned} w_{r,t} &= \frac{e_{a,t} + \theta_{a,t}(\theta_{a,t} - \theta_{r,t})}{e_{r,t} + e_{a,t} + (\theta_{r,t} - \theta_{a,t})^2} \\ w_{a,t} &= 1 - w_{r,t} \end{aligned} \tag{5.20}$$

During the first time period, the simple average is used. In subsequent time periods, the biases θ, which is the sum of intentional and unintentional biases, up to time period t is calculated as:

$$\theta_{i,t} = \frac{1}{t-1-18} \sum_{n=19}^{t-1} (q_{i,n} - d_n) \tag{5.21}$$

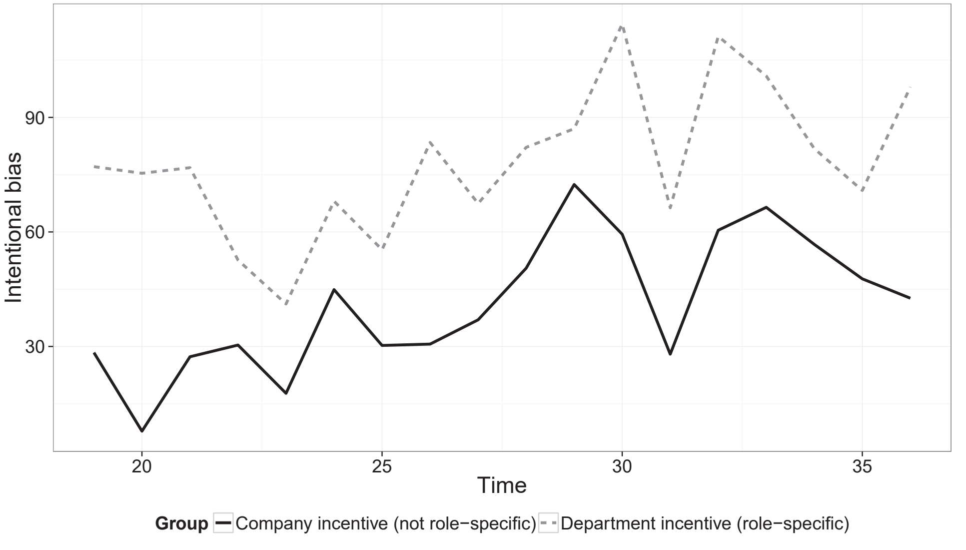
Performance under this de-biasing scheme is compared to the performance under the other weighing schemes, using RMSE. A Wilcoxon signed-rank test is used for hypothesis (6) to determine whether this scheme outperforms weighing schemes without such a correction. The simple average is excluded from the comparison, because it is used in the first phase of the experiment, which has a different time series of simulated demand.

5.4 Results

We first examine the average intentional bias of participants of the experiment for the two types of incentives. We then proceed by examining the results of our estimated statistical model (5.18) and statistical tests, and discuss the implications of these results for our hypotheses in turn.

Figure 5.1 shows the mean intentional bias $\delta_{i,t}$ aggregated per time period over participants, for the department and company incentive separately, and with the simple average as the weighing scheme. The biases are positive under both incentives, and the intentional bias for the role-specific department incentive is consistently and substantially higher than the bias for the company incentive. Roles, even without role-specific incentives, seem to entail intentional biases, amounting to an average adjustment of 41.032. In addition, role-specific department incentives almost double the intentional biases, from 41.032 to 78.336.

Figure 5.1: **Mean intentional biases**
This figure shows the mean intentional bias δ_t aggregated over participants (students and practitioners combined) per time period, for the department and company incentive with the simple average as the weighing scheme.



To examine the intentional biases δ in more detail, Table 5.2 presents the estimates of the AR(1) model (5.18). The estimated effect of the role-specific department incentive β is substantially and significantly different from zero for both students ($\beta =$

Table 5.2: **Estimates of the AR(1) model for intentional adjustments**
Estimates of model (5.18) for the groups of students and practitioners separately. The effect of the department incentive β is significant. Standard errors are in parentheses.

	Students (n = 357)	Practitioners (n = 72)
μ_α	32.972 (0.122)	33.824 (0.756)
μ_ϕ	0.298 (0.079)	0.249 (0.655)
β	24.826 (0.003)	25.370 (0.721)
σ^2_α	8.661 (0.978)	12.685 (0.431)
σ^2_ϕ	0.091 (0.266)	0.143 (0.734)

$24.826, \text{s.e.} = 0.003$) and practitioners ($\beta = 25.370, \text{s.e.} = 0.721$), which indicates that role-specific department incentives increase intentional biases, consistent with hypothesis 2. The mean (5.19) of the AR(1) process is equal to 46.968 for intentional biases under company incentives and 82.333 for intentional biases under role-specific department incentives. As the average forecast is equal to 572.856, these biases correspond to a 8% and 14% adjustment. The intentional bias under the company incentive shows that roles, even without role-specific incentives, entail intentional biases, which supports hypothesis 1.

Table 5.3: **Estimates of the AR(1) model for different types of forecasting behavior**
Estimates for different subsets of the data, either based on all participants combined, or separately for chasers and smoothers. Standard errors are in parentheses. The estimates for chasers and smoothers are not significantly different.

	All (n = 357)	Chasers (n = 271)	Smoothers (n = 86)
μ_α	32.972 (0.122)	34.344 (0.099)	33.620 (0.541)
μ_ϕ	0.298 (0.079)	0.387 (0.081)	0.351 (0.575)
β	24.826 (0.003)	28.258 (0.025)	29.857 (0.763)
σ^2_α	8.661 (0.978)	6.996 (0.623)	5.027 (12.264)
σ^2_ϕ	0.091 (0.266)	0.084 (0.454)	0.124 (0.864)

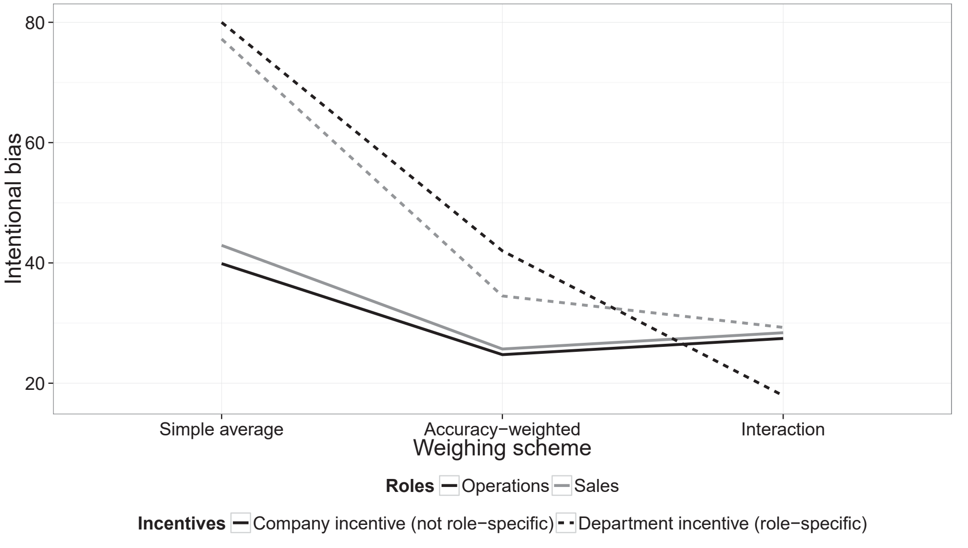
Estimating the model separately for the two types of forecasting behavior introduced in Chapter 4 gives the results in Table 5.3. The differences are not statistically significant, implying that intentional biases, like unintentional biases, are unrelated to the two types, chasers and smoothers, of forecasting behavior, which supports hypothesis 3.

To explore how intentional biases change under the various weighing schemes, Table 5.4 lists the average descriptive intentional biases per role and incentive for each of the weighing schemes. Figure 5.2 illustrates these outcomes. Under the simple average scheme, the biases for the company incentive are substantial, 39.886 (s. e. 5.731) for operations and 42.921 (s. e. 3.192) for sales. This shows that roles, even without role-specific financial incentives, entail intentional biases, supporting hypothesis 1. The bias under the department incentive is significantly larger, 79.998 (s. e. 3.654) for operations and 77.229 (s. e. 3.685) for sales. This shows that financial role-specific incentives result in larger intentional biases, supporting hypothesis 2. The accuracy-weighted scheme approximately halves the intentional biases found under the simple average: the smallest drop is 38% for operations under a company incentive from 39.886 to 24.751, and the largest drop is 55% for sales under a department incentive from 77.229 to 34.507. The accuracy-weighted scheme reduces, but does not remove, the intentional bias, as the lowest bias, 24.751 for operations under a company incentive, is still substantial. Thus, it removes neither the intentional bias due to role-specific incentives nor the bias due to roles themselves, which supports hypothesis 4. Including interaction between the participant and agent, allowing them to revise their inputs, lowers the intentional biases to a similar extent as the accuracy-weighted scheme does, though with more variation: the smallest drop is 31% for operations under a company incentive from 39.886 to 27.442, and the largest drop is 78% for operations under a department incentive from 79.998 to 17.942. This contrasts with hypothesis 5, which posits that interactions emphasize the competitive nature of the task. Rather, agents' revision of forecasts towards the quantities provided by participants appear to stimulate cooperation, reducing participants' intentional bias.

Table 5.4: **Mean intentional biases under the various weighing schemes**
The columns first differentiate between department and company incentive, followed by the role of the participant. The rows differentiate between the weighing schemes. Standard errors in parentheses. Differences between incentives within weighing schemes and between the simple average and alternative weighing schemes are significant at $p < 0.05$.

	Company incentive (Not role-specific)		Department incentive (Role-specific)	
	Operations	Sales	Operations	Sales
Simple average	39.89 (5.73)	42.92 (3.19)	79.99 (3.65)	77.23 (3.69)
Accuracy-weighted	24.75 (4.52)	25.69 (2.56)	42.00 (3.45)	34.51 (3.10)
Interaction	27.44 (9.95)	28.39 (5.69)	17.94 (7.52)	29.28 (7.14)

Figure 5.2: **Average intentional bias for roles, incentives and weighing Schemes**
This figure shows the descriptive intentional biases from Table 5.4.



The weighing schemes—the simple average (5.14), the accuracy-weighted (5.15), and interaction between participant and agent—influence forecasting behavior and directly affect the size of the intentional biases. However, none of the schemes fully removes the intentional biases. Moreover, unintentional biases are also present in the forecast. Table 5.5 lists the accuracy under the accuracy-weighted scheme, the interaction scheme, and the de-biasing scheme (5.20). The de-biasing scheme greatly reduces inaccuracy by 35% compared to the accuracy-weighted scheme, supporting hypothesis 6. Compared to the interaction scheme it reduces inaccuracy by 37%. Even such a basic adjustment, which estimates and removes a constant bias from the inputs of the participant, gives considerable gains.

Table 5.5: **Accuracy**
Accuracy under the various weighing schemes. Standard errors in parentheses.

	RMSE
Accuracy-weighted	197.163 (1.821)
Interaction	203.934 (3.547)
De-biasing scheme	128.097 (1.593)

5.5 Discussion and conclusion

By conducting an elaborate experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, with a large group of students and practitioners, this study has examined intentional biases in the context of different departmental roles and incentives in the organizational forecasting process. We evaluated the effects of roles, incentives, and various weighing schemes on behavior and performance and find that roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We test the claim that an accuracy-weighted scheme can remove unintentional biases, and conclude that though it can half these biases, it cannot fully negate them. Finally, we observe that a simple de-biasing scheme shows great promise in reducing intentional as well as unintentional biases by 35%.

These contributions are important for research into forecast biases by isolating intentional biases for study. The study extends work by Kuo and Liang (2004), Önkal et al. (2012), and Yaniv (2011), which does not distinguish between intentional and unintentional biases. Moreover, by differentiating between unintentional and intentional biases, and by studying behavior under a weighing scheme, our experiment mimics the case study of Oliva and Watson (2009) and Oliva and Watson (2011). Not only do we provide empirical support for the impact of weighing schemes, but we also show that these schemes do not entirely remove intentional biases. More importantly, we determine the limits of the accuracy-weighted scheme, and show that alternatives, such as a simple de-biasing scheme, have a potential for a much larger gain.

Our behavioral experiment is the first to isolate intentional biases and to assess how they are affected by weighing schemes. Future work can build on this by exploring additional decision-making contexts. Other weighing schemes can be used, and the interaction between participants and agents can be extended. In addition, more roles, such as marketing and finance, can be included in the experiment, which removes the simple dichotomy of sales and operations used in our experiment, and allows for more diverse roles. Similarly, other and more elaborate incentive schemes can be introduced.

Our current insights already have important consequences for the design of the forecasting process in terms of coordination mechanisms and incentives (Singhal and Singhal, 2007). This is important for practice, because forecasters' behavior directly affects forecast performance, which can have large financial ramifications (Fildes et al., 2008; Syntetos et al., 2011, 2010). Our work on disentangling specific design choices and examining these in isolation paves the way for future work on forecast process design, and specifically the potential performance gain of weighing schemes.

More immediately, however, it presses for a careful review of current policies, because choices in terms of roles, incentives, and weighing schemes meant to increase performance can have a detrimental effect.

Chapter 6

Summary and Conclusion

The demand that drives various activities in the supply chain is inherently uncertain, necessitating the need for forecasting. Retailers require forecasts for sales, inventory and order decisions, suppliers for production and procurement decisions, and distributors for capacity allocation decisions. In practice, forecast errors are substantial, which negatively affects operational performance (Danese and Kalchschmidt, 2011; Enns, 2002; Hughes, 2001; Ritzman and King, 1993; Zhao and Xie, 2002). Reducing or minimizing these forecast errors is central to this thesis and is achieved by improving the forecasting capabilities of companies, which encompasses both extending the available forecasting methods and models as well as analyzing how the forecasting process, the context in which these methods and models are embedded, can be improved.

Extending the available forecasting methods and models is done in two different studies, which can be summarized as exploiting already available information, without context-specific assumptions, to achieve substantial gains. The first study exploits the available information by generalizing existing forecasting methods to better model intermittent demand. Existing methods either ignore a dependency between the time between orders and order size, or focus on the risk of inventory obsolescence. This limited scope is costly in terms of inventory and financial performance. The second study also generalizes existing forecasting methods and supersedes the traditional discussion of top-down versus bottom-up methods, by examining how the hierarchy of products is used in forecasting. Stock-keeping units (SKUs) naturally group together in a hierarchy going from the bottom, with individual sales per product, through several intermediary levels, denoting sales for groups of related products at increasingly general aggregation levels, such as product groups and categories, to the top of the hierarchy, which lists total sales. Two commonly used approaches in practice and research start from opposite ends of the hierarchy to generate forecasts for all series: bottom-up forecasting and top-down forecasting. Both of these approaches imply a loss of information because the scope is restricted to separate and independent initial forecasts. However, by generating joint forecasts for a group of products directly, information is better used which translates to superior performance. This approach explicitly incorporates product dependencies, such as the complementarity of products and product substitution, which are otherwise ignored. Whereas the first study exploits available information for each SKU separately, the second study does so by considering SKUs in groups and hierarchies, expanding the scope of the forecasting models.

To complement the application of forecasting methods and models in the first two studies, the third and fourth studies analyze the forecasting process at companies, and specifically the use of judgment. Judgmental forecasting is central to the forecasting processes at many companies, and directly affects supply chain performance (Fildes et al., 2008; Syntetos et al., 2011, 2010). These studies provide insights into how forecaster behavior systematically differs and how this and the design of the forecasting process affect performance. We demonstrate that forecasting behavior differs systematically between individuals to the extent that we discern two markedly different types of forecasters, labeled chasers and smoothers. We also examine the influence of roles and incentives, and trace the extent to which forecasters intentionally adjust their forecasts, and how this is affected by design choices.

The following sections summarize the main findings of the four specific studies, discuss scientific contributions and managerial implications, and provide suggestions for future research.

6.1 Main findings

Chapter 2 presents an intermittent demand forecasting method that conditions on the elapsed time since the last demand occurrence to anticipate incoming demand and shows, using empirical data, that this can substantially reduce both stock investment and lost revenue for spare parts. We extensively benchmark our method against existing forecasting and bootstrapping methods on forecast accuracy and inventory performance and demonstrate that its performance is robust under general conditions. Existing forecasting methods either do not change the forecast after a period of zero demand, ignoring all forms of cross-correlations, or adjust the forecast downwards, addressing only the specific case of inventory obsolescence and not the general forms of cross-correlations observed in empirical data. All methods ignore the fact that activities at the source of the demand, such as aggregation of demand, preventive and corrective maintenance, can lead to a positive relation between demand size and inter-arrival time of demand occurrences. By anticipating incoming demand, our method offers substantial financial gains.

Chapter 3 looks into generating forecasts for product groups, and specifically examines product dependencies ignored in practice. Forecasts are often made at various levels of aggregation of individual products, which combine into groups at higher hierarchical levels. We provide an alternative to the traditional discussion of bottom-up versus top-down forecasting by examining how the hierarchy of products can be exploited when forecasts are generated. Instead of selecting series from parts of the hierarchy for forecasting, we explore using all the series. Moreover, instead of using the hierarchy after initial forecasts are generated, we consider the hierarchical series as a whole to instantaneously generate forecasts for all levels of the hierarchy. Our integrated approach explicitly incorporates product dependencies, such as complementarity of products and product substitution, which are otherwise ignored. A simulation study, comparing and contrasting existing approaches from literature under possible cross-correlations and dependencies, shows the conditions under which an integrated approach is advantageous. An empirical study shows the substantial gain, in terms of forecast performance as well as inventory performance, of generalizing the bottom-up and top-down forecasting approaches to an integrated approach. The integrated approach is applicable to hierarchical forecasting in general, and extends beyond the current application of forecasting for manufacturers.

Chapter 4 demonstrates that forecasting behavior differs systematically between individuals to the extent that we discern two markedly different types of forecasters. One is characterized by overreaction to forecast errors and might be labeled chasers, while the other is characterized by underreaction to forecast errors, and might be labeled smoothers. Extending the models used in earlier behavioral experiments, our approach relies on wavelets and state space modeling to incorporate forecasting heterogeneity. We demonstrate that contextual biases can only be meaningfully explored after controlling for the forecaster's inclination towards chasing or smoothing. We further show that departmental biases persistently impact judgmental forecasting, even if forecasts are constructed to be free of intentional biases.

Chapter 5 examines intentional biases, an overlooked research area, that arise due to the influence of different departmental roles and incentives in the forecasting process. Through an experiment, which simulates forecasting and production quantity decisions in an interdepartmental decision-making context, we examine the effects of roles, incentives, and various weighing schemes on behavior and performance. We find that roles, even without role-specific incentives, entail intentional biases of 8% of the forecast, and that role-specific incentives increase these biases to 14%. We test the claim that an accuracy-weighted scheme can remove unintentional biases, and conclude that though this halves these biases, it does not fully remove them. Finally, we observe that a weighing scheme that explicitly corrects biased inputs shows great promise in reducing intentional as well as unintentional biases. In our experiment, this scheme reduces biases by 35%.

6.2 Scientific contributions

The four studies generally show that the conflation of forecast information and forecasting capability, used in the stylized models of Aviv (2001) and Aviv (2007), is an unwarranted simplification. Limitations in the capability of retailers and manufacturers, in terms of forecast model formulation and estimation (Småros, 2007), but also in terms of the design of the forecasting process, are a concern. Hence, the studies in this thesis analyze and draw conclusions based on empirical data collected from industry.

Because of limitations in forecasting capability, not all of the already available information is used by companies, which suggests that simply expanding the forecast information available is futile. Chapter 2 demonstrates that, even without context-specific knowledge and assumptions, and even if there is very little information available, currently available information can still be used to substantially improve performance. By extending forecasting methods from literature, both parametric (Croston, 1972; Snyder et al., 2012; Syntetos and Boylan, 2001, 2005, 2006; Syntetos et al., 2012; Teunter et al., 2011), and nonparametric (Willemain et al., 2004), the forecasting capability of manufacturers is directly increased. This chapter is the first to propose a forecasting method that can accommodate a positive relation between demand size and inter-arrival time of demand occurrences, which possibly arises due to activities at the source of the demand, such as aggregation of demand, preventive and corrective maintenance (Altay et al., 2012; Boylan and Syntetos, 2007; Wang and Syntetos, 2011; Willemain et al., 1994). The approach extends the literature by specifically examining the overlooked case of positive cross-correlation between the demand size and the elapsed time, and has shown its importance: only focusing on obsolescence comes at a cost. All previously existing forecasting methods either ignore all forms of cross-correlations or address only the specific case of inventory obsolescence. Moreover, this chapter extensively benchmarks these methods on several data sets on forecast accuracy and inventory performance.

While Chapter 2 treats products independently, as the little available information is not enough to estimate dependencies between products, Chapter 3 looks into generating forecasts for product groups, and specifically examines product dependencies ignored in practice. Two commonly used approaches in practice and research start from opposite ends of the hierarchy to generate forecasts for all series: bottom-up forecasting and top-down forecasting (Widiarta et al., 2009). In bottom-up forecasting, base forecasts are generated for product demand at the lowest level in the hierarchy (Gordon et al., 1997). Subsequently, these are aggregated to determine forecasts at higher hierarchical levels. Bottom-up forecasting is commonly contrasted with top-down forecasting, in which forecasts are generated for aggregated demand and disaggregated downwards to determine forecasts at lower levels in the hierarchy (Kahn, 1998). Research stretches over three decades with mixed results as to preference for either bottom-up or top-down forecast approaches. The integrated approach supersedes the traditional comparison of bottom-up and top-down approaches (Fliedner, 1999; Kahn, 1998), by generating forecasts at all hierarchical levels and incorporating all available information, rather than only using selected parts of available data. The integrated approach avoids ex-post revising of forecasts, as is done in the combination approach (Hyndman et al., 2011), as generated forecasts are already reconciled and respect the additive restrictions placed on the series by the hierarchy.

Chapters 4 and 5 heed the call that research has to extend beyond the technical side of forecast generation and consider how the forecasting process is managed and organized (Armstrong, 1987; Danese and Kalchschmidt, 2011). There is a lack of performance evaluation and management of forecasting processes at companies, and a blurred distinction between forecasts, plans, and goals (Moon et al., 2003). Moreover, forecasts generated by forecasting methods are not directly used. The use of judgment for generating and adjusting forecasts is often preferred and widely used (Hughes, 2001; Lawrence et al., 2000).

The contribution of Chapter 4 lies in the assessment of the consequences of heterogeneity for judgmental forecasting, and our findings have important repercussions for theory building based on evidence derived from aggregate results. A major problem of the current knowledge on judgmental biases and the performance of judgmental forecasting is that most of the evidence on is at an aggregate level, encompassing large groups of individuals (in the case of experiments based on the newsvendor model see e.g. Bolton and Katok, 2008; Bostian et al., 2008; Kremer et al., 2011; Schweitzer and Cachon, 2000). This is problematic, because it overlooks the existence and impact of forecasting heterogeneity, which refers to the possibility that forecasting behavior differs systematically between individuals. It may well be the case that two types of forecasters differ in the extent to which they overreact or underreact to forecasting errors, and display chasing or smoothing behavior. Such heterogeneity of individual biases possibly leads to inaccurate aggregate results, which do not reflect individual behavior (Lau et al., 2014). Chapter 4 extends earlier behavioral experiments of Bostian et al. (2008), Kremer et al. (2011), and Schweitzer and Cachon (2000) to demonstrate, using an approach relying on wavelets and state space modeling, that forecasting behavior indeed differs systematically between individuals. That is, forecasters can be divided into people who overreact to forecast errors and display chasing behavior, and people who underreact to forecast errors, and thus display smoothing behavior.

The existence of different types of forecasting behavior leads to the conclusion that forecasters overreact in a relatively unstable environment, which conflicts with the conclusion of Kremer et al. (2011). The difference between chasers and smoothers and their behavior also explains why we find no evidence for learning effects in contrast to Bostian et al. (2008). Furthermore, Chapter 4 demonstrates the existence of persistent departmental biases of roles and incentives. In line with conclusions of Kuo and Liang (2004) and Önkal et al. (2012), we find that forecasting behavior is influenced by roles. In contrast to the conclusion of Yaniv (2011), the effect of roles is not negated using incentives. We are unable to differentiate between intentional and unintentional biases, as roles have a strong effect, even without incentives, which has ramifications as we can no longer assume that we can disentangle the two biases (e.g. Oliva and Watson, 2009, 2011).

Chapter 5 focuses on intentional biases, an overlooked research area. Biases in judgmental forecasting have often been studied, but unintentional and intentional biases have never been disentangled. By isolating intentional biases for study, this chapter extends the work by Kuo and Liang (2004), Önkal et al. (2012), and Yaniv (2011), which do not distinguish between intentional and unintentional biases. Our work shows the importance of disentangling the two sources of biases for research, as intentional biases are substantial and present even without financial incentives.

Moreover, by differentiating between unintentional and intentional biases, and by studying behavior under a weighing scheme, our experiment in Chapter 5 mimics the case study of Oliva and Watson (2009) and Oliva and Watson (2011). Not only do we provide empirical support for the impact of weighing schemes, but we also show that these schemes do not entirely remove intentional biases. More importantly, we determine the limits of the accuracy-weighted scheme, and show that alternatives, such as a simple-debiasing scheme, have a potential for a much larger gain.

6.3 Managerial implications

Generally, companies lack the knowledge, expertise and training in the field of forecasting to validly support decision-making (Hughes, 2001). The situation has even become worse, as the level of knowledge and forecast accuracy have decreased over time (McCarthy et al., 2006). Davis and Mentzer (2007) observe a gap between theory and practice in terms of forecasting capability, and consider this a significant issue. Our studies introduce models and methods that directly extend forecasting capability, and provide insights into how design choices of the forecasting process affect behavior and performance.

A managerial implication of Chapter 2 is that the nature of the demand process is important and has to be considered for forecasting and inventory decisions. Exclusively focusing on the risk of obsolescence leads to much higher costs, as possible decisions are only taken over a restricted domain. A specific managerial implication of this chapter is that we derive an easy to implement and novel method that can immediately be used for inventory decisions for SKUs, even if context-specific knowledge is unavailable. We also provide the means to assess to which SKUs this method should be applied for the largest gain. This also allows managers to apply this method on a smaller scale and facilitates implementation. The analysis of financial performance shows the importance of applying our method. Our method gave the largest reduction in inventory investment of 14% and even reduced lost revenue by 4%, thus clearly outperforming all other methods. It is easy to estimate and proves to be robust in a range of applications, and is thus generally, and immediately, applicable in practice.

The integrated approach of Chapter 3 is applicable to hierarchical forecasting in general, and extends beyond the current application of forecasting for manufacturers. Even overlapping groups of products can be easily accommodated. The large reductions in stock investments, up to as much as a 39%, show that the forecast performance directly translates to large financial gains, and is highly relevant for forecasting processes at companies. The simulation study, which compares and contrasts existing approaches under possible cross-correlations and dependencies, demonstrates under which conditions our integrated approach is advantageous. Furthermore, our empirical study shows the substantial gain, in terms of forecasting performance as well as inventory performance, of generalizing the bottom-up and top-down forecast approaches to an integrated approach. All available information is used, product dependencies are taken into account, such as the complementarity of products and product substitution, and other features of the series are incorporated as well, such as seasonality, which are otherwise ignored. Additional advantages of formulating the integrated approach as a state space model are that outliers, missing values, and extra information, such as pertaining to promotions, can be easily, and flexibly, included (Durbin and Koopman, 2012; Harvey, 1989).

The findings of Chapter 4 are important for practice because forecast behavior directly affects forecast performance, which can have large financial ramifications.

Chasers and smoothers have substantially different forecast performance, so that recognizing the difference between these two types of forecasting can lead to better hiring and training practices for forecasters. The impact of departmental biases also has ramifications for how the forecasting process is orchestrated within companies when multiple departments participate, as roles and the behavior of other participants affect behavior.

The insights of Chapter 5 have important ramifications for the design of the forecasting process in terms of coordination mechanisms and incentives by quantifying the impact of roles and incentives (Singhal and Singhal, 2007). This is important for practice, because forecasters' behavior directly affects forecast performance, which can have large financial ramifications (Fildes et al., 2008; Syntetos et al., 2011, 2010). This chapter presses for a careful review of current policies and practices for the forecasting processes at companies.

6.4 Future research

The specific case examined in Chapter 2 is only one implementation of the general method described. More research is needed to explore the dependency between demand size and elapsed time on empirical data sets, but also to apply more general models, which more broadly incorporate the dynamics between demand size and elapsed time. These dynamics especially come into play due to the product life cycle, so that it can become important to not only classify SKUs once but foresee how the characteristics evolve over time, so that they can be incorporated in the method. Our method could also be suited for applications outside of spare parts management.

For Chapter 3, future work can extend the estimation part of elaborate state space models. A drawback of the integrated approach is that it is computationally more demanding than the other approaches, but not to such an extent that it bars use in conventional software used by manufacturers. For large numbers of products, principal component analysis can be used, but much work remains to be done for the efficient estimation of large state space models.

Chapters 4 and 5 show the need for future work on behavioral experiments. Different types of forecasting behavior will remain an important topic for future study, as they impact both research done so far and practice. The novel methodology we outlined in Chapter 4, relying on wavelets and state space modeling, should prove to be flexible in similar types of research. The behavioral experiment in Chapter 5 is the first to isolate intentional biases and to assess how they are affected by weighing schemes. Future work can build on this by exploring additional decision-making contexts. Other weighing schemes can be used, and the interaction between participants and agents can be extended. In addition, more roles, such as marketing and finance, can be included in the experiment, replacing the simple dyad of sales and operations used in our experiment, and allowing for more diverse roles. Similarly, other and more elaborate incentive schemes can be introduced. Our work on disentangling specific design choices and examining these in isolation paves the way for future work on forecast process design, and specifically the potential performance gain of weighing schemes.

In general, future work can extend analyses of all four studies to accommodate information and managers from other parties in the supply chain, such as retailers and suppliers, to further improve the forecasting capabilities of companies.

References

Abramovich, F., T. C. Bailey, T. Sapatinas. 2000. Wavelet analysis and its statistical applications. Journal of the Royal Statistical Society: Series D (The Statistician), 49 (1), 1–29.

Ali, M. M., J. E. Boylan. 2011. Feasibility principles for downstream demand inference in supply chains. Journal of the Operational Research Society, 62 (3), 474–482.

Ali, M. M., J. E. Boylan, A. A. Syntetos. 2012. Forecast errors and inventory performance under forecast information sharing. International Journal of Forecasting, 28 (4), 830–841.

Altay, N., L. A. Litteral, F. Rudisill. 2012. Effects of correlation on intermittent demand forecasting and stock control. International Journal of Production Economics, 135 (1), 275–283.

Anandalingam, G., L. Chen. 1989. Bayesian forecast combination and Kalman filtering. International Journal of Systems Science, 20 (8), 1499–1507.

Aretz, K., S. M. Bartram, P. F. Pope. 2011. Asymmetric loss functions and the rationality of expected stock returns. International Journal of Forecasting, 27 (2), 413–437.

Armstrong, J. S. 1987. The handbook of forecasting: the forecasting audit. S. Makridakis, S. C. Wheelwright, eds., The Handbook of Forecasting. John Wiley, New York, 584–602.

Arthur, D., S. Vassilvitskii. 2007. K-means++: the advantages of careful seeding. Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, 1027–1035.

Ashiya, M. 2009. Strategic bias and professional affiliations of macroeconomic forecasters. Journal of Forecasting, 28 (2), 120–130.

Ashton, A. H., R. H. Ashton. 1985. Aggregating subjective forecasts: some empirical results. Management Science, 31 (12), 1499–1508.

Athanasopoulos, G., R. A. Ahmed, R. J. Hyndman. 2009. Hierarchical forecasts for Australian domestic tourism. International Journal of Forecasting, 25 (1), 146–166.

Aviv, Y. 2001. The effect of collaborative forecasting on supply chain performance. Management Science, 47 (10), 1326–1343.

Aviv, Y. 2002. Gaining benefits from joint forecasting and replenishment processes: the case of auto-correlated demand. Manufacturing & Service Operations Management, 4 (1), 55–74.

Aviv, Y. 2007. On the benefits of collaborative forecasting partnerships between retailers and manufacturers. Management Science, 53 (5), 777–794.

Babai, M. Z., M. M. Ali, J. E. Boylan, A. A. Syntetos. 2013. Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: theory and empirical analysis. International Journal of Production Economics, 143 (2), 463–471.

Babai, M. Z., A. A. Syntetos, R. H. Teunter. 2014. Intermittent demand forecasting: an empirical study on accuracy and the risk of obsolescence. International Journal of Production Economics, (157), 212–219.

Bates, J. M., C. W. J. Granger. 1969. The combination of forecasts. OR, 20 (4), 451.

Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, L. A. Thorsrud. 2012. Does forecast combination improve Norges bank inflation forecasts? Oxford Bulletin of Economics and Statistics, 74 (2), 163–179.

Bolton, G. E., E. Katok. 2008. Learning by doing in the newsvendor problem: a laboratory investigation of the role of experience and feedback. Manufacturing & Service Operations Management, 10 (3), 519–538.

Bolton, G. E., A. Ockenfels, U. W. Thonemann. 2012. Managers and students as newsvendors. Management Science, 58 (12), 2225–2233.

Bostian, A. A., C. A. Holt, A. M. Smith. 2008. Newsvendor 'pull-to-center' effect: adaptive learning in a laboratory experiment. Manufacturing & Service Operations Management, 10 (4), 590–608.

Bourland, K. E., S. G. Powell, D. F. Pyke. 1996. Exploiting timely demand information to reduce inventories. European Journal of Operational Research, 92 (2), 239–253.

Boylan, J. E., A. A. Syntetos. 2007. The accuracy of a modified Croston procedure. International Journal of Production Economics, 107 (2), 511–517.

Boylan, J. E., A. A. Syntetos, G. C. Karakostas. 2008. Classification for forecasting and stock control: a case study. Journal of the Operational Research Society, 59 (4), 473–481.

Brockhoff, K. 1983. Group processes for forecasting. European Journal of Operational Research, 13 (2), 115–127.

Byron, R. P. 1978. The estimation of large social account matrices. Journal of the Royal Statistical Society. Series A (General), 141 (3), 359.

Cachon, G. P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. Management Science, 46 (8), 1032–1048.

Cantor, D. E., J. R. Macdonald. 2009. Decision-making in the supply chain: examining problem solving approaches and information availability. Journal of Operations Management, 27 (3), 220–232.

Chan, Y. L., J. H. Stock, M. W. Watson. 1999. A dynamic factor model framework for forecast combination. Spanish Economic Review, 1 (2), 91–121.

Chaovalit, P., A. Gangopadhyay, G. Karabatis, Z. Chen. 2011. Discrete wavelet transform-based time series analysis and mining. ACM Computing Surveys (CSUR), 43 (2), 6.

Chatfield, C. 1996. Model uncertainty and forecast accuracy. Journal of Forecasting, 15 (7), 495–508.

Chen, F., Z. Drezner, J. K. Ryan, D. Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting, lead times, and information. Management Science, 46 (3), 436–443.

Chen, L., H. L. Lee. 2009. Information sharing and order variability control under a generalized demand model. Management Science, 55 (5), 781–797.

Clemen, R. T. 1989. Combining forecasts: a review and annotated bibliography. International Journal of Forecasting, 5 (4), 559–583.

Clemen, R. T., R. L. Winkler. 1986. Combining economic forecasts. Journal of Business & Economic Statistics, 4 (1), 39–46.

Clemen, R. T., R. L. Winkler. 1999. Combining probability distributions from experts in risk analysis. Risk Analysis, 19 (2), 187–203.

Cohen, M. A., H. L. Lee. 1990. Out of touch with customer needs? Spare parts and after sales service. Sloan Management Review, 31 (2), 55–66.

Costantini, M., R. M. Kunst. 2011. Combining forecasts based on multiple encompassing tests in a macroeconomic core system. Journal of Forecasting, 30 (6), 579–596.

Croston, J. D. 1972. Forecasting and stock control for intermittent demands. Operational Research Quarterly, 23 (3), 289–303.

Crowley, P. M. 2007. A guide to wavelets for economists. Journal of Economic Surveys, 21 (2), 207–267.

Danese, P., M Kalchschmidt. 2011. The role of the forecasting process in improving forecast accuracy and operational performance. International Journal of Production Economics, (131), 204–214.

Dangerfield, B. J., J. S. Morris. 1992. Top-down or bottom-up: aggregate versus disaggregate extrapolations. International Journal of Forecasting, 8 (2), 233–241.

Davis, D. F., J. T. Mentzer. 2007. Organizational factors in sales forecasting management. International Journal of Forecasting, 23 (3), 475–495.

De Véricourt, F., K. Jain, J. N. Bearden, A. Filipowicz. 2013. Sex, risk and the newsvendor. Journal of Operations Management, 31 (1), 86–92.

Diebold, F. X., P. Pauly. 1987. Structural change and the combination of forecasts. Journal of Forecasting, 6 (1), 21–40.

Diebold, F. X., P. Pauly. 1990. The use of prior information in forecast combination. International Journal of Forecasting, 6 (4), 503–508.

Durbin, J., S. J. Koopman. 2012. Time series analysis by state space methods. Oxford University Press.

Eaves, A. H. C., B. G. Kingsman. 2004. Forecasting for the ordering and stock-holding of spare parts. Journal of the Operational Research Society, 55 (4), 431–437.

Edwards, J. B., G. H. Orcutt. 1969. Should aggregation prior to estimation be the rule? The Review of Economics and Statistics, 51 (4), 409–420.

Engel, E. M. R. A. 1984. A unified approach to the study of sums, products, time-aggregation and other functions of ARMA processes. Journal of Time Series Analysis, 5 (3), 159–171.

Enns, S. T. 2002. MRP performance effects due to forecast bias and demand uncertainty. European Journal of Operational Research, 138 (1), 87–102.

Fang, Y. 2003. Forecasting combination and encompassing tests. International Journal of Forecasting, 19 (1), 87–94.

Fildes, R., P. Goodwin, M. Lawrence, K. Nikolopoulos. 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting, 25 (1), 3–23.

Fildes, R., K. Nikolopoulos, S. F. Crone, A. A. Syntetos. 2008. Forecasting and operational research: a review. Journal of the Operational Research Society, 59 (9), 1150–1172.

Fliedner, E. B., V. A. Mabert. 1992. Constrained forecasting: some implementation guidelines. Decision Sciences, 23 (5), 1143.

Fliedner, G. 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. Computers & Operations Research, 26 (10), 1133–1149.

Fliedner, G. 2001. Hierarchical forecasting: issues and use guidelines. Industrial Management & Data Systems, 101 (1), 5–12.

Gardner, E. S. 1990. Evaluating forecast performance in an inventory control system. Management Science, 36 (4), 490–499.

Gardner, E. S., A. B. Koehler. 2005. Comments on a patented bootstrapping method for forecasting intermittent demand. International Journal of Forecasting, 21 (3), 617–618.

Gaur, V., A. Giloni, S. Seshadri. 2005. Information sharing in a supply chain under ARMA demand. Management Science, 51 (6), 961–969.

Gavrilov, M., D. Anguelov, P. Indyk, R. Motwani. 2000. Mining the stock market: which measure is best? Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 487–496.

Gençay, R., F. Selçuk, B. J. Whitcher. 2001. An introduction to wavelets and other filtering methods in finance and economics. Academic Press.

Gilbert, K. 2005. An ARIMA supply chain model. Management Science, 51 (2), 305–310.

Gino, F., G. Pisano. 2008. Toward a theory of behavioral operations. Manufacturing & Service Operations Management, 10 (4), 676–691.

Glueck, J. J., P. Koudal, W. Vaessen. 2011. The service revolution: manufacturing's missing crown jewel. Research report. Deloitte Development LLC.

Gönül, S., D. Önkal, P. Goodwin. 2009. Expectations, use and judgmental adjustment of external financial and economic forecasts: an empirical investigation. Journal of Forecasting, 28 (1), 19–37.

Goodwin, P., R. Fildes. 1999. Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? Journal of Behavioral Decision Making, 12 (1), 37–53.

Goodwin, P., R. Fildes, M. Lawrence, K. Nikolopoulos. 2007. The process of using a forecasting support system. International Journal of Forecasting, 23 (3), 391–404.

Goodwin, P., G. Wright. 1993. Improving judgmental time series forecasting: a review of the guidance provided by research. International Journal of Forecasting, 9 (2), 147–161.

Gordon, T. P., J. S. Morris, B. J. Dangerfield. 1997. Top-down or bottom-up: which is the best approach to forecasting? The Journal of Business Forecasting Methods & Systems, 16 (3), 13–16.

Graefe, A., J. S. Armstrong. 2011. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. International Journal of Forecasting, 27 (1), 183–195.

Granger, C. W. J., R. Ramanathan. 1984. Improved methods of combining forecasts. Journal of Forecasting, 3 (2), 197–204.

Gross, C. W., J. E. Sohl. 1990. Disaggregation methods to expedite product line forecasting. Journal of Forecasting, 9 (3), 233–254.

Ha, A. Y., S. Tong. 2008. Contracting and information sharing under supply chain competition. Management Science, 54 (4), 701–715.

Hamilton, J. D. 1994. Time series analysis. Princeton University Press.

Harvey, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Hogarth, R. M., S. Makridakis. 1981. Forecasting and planning: an evaluation. Management Science, 27 (2), 115–138.

Huang, G. Q., Jason S. K. Lau, K. L. Mak. 2003. The impacts of sharing production information on supply chain dynamics: a review of the literature. International Journal of Production Research, 41 (7), 1483–1517.

Hughes, M. C. 2001. Forecasting practice: organisational issues. Journal of the Operational Research Society, 52 (2), 143–149.

Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, H. L. Shang. 2011. Optimal combination forecasts for hierarchical time series. Computational Statistics & Data Analysis, 55 (9), 2579–2589.

Hyndman, R. J., A. B. Koehler. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting, 22 (4), 679–688.

Inderfurth, K., R. Kleber. 2013. An advanced heuristic for multiple-option spare parts procurement after end-of-production. Production and Operations Management, 22 (1), 54–70.

Iyer, A. V., J. Ye. 2000. Assessing the value of information sharing in a promotional retail environment. Manufacturing & Service Operations Management, 2 (2), 128–143.

Johnston, F. R., J. E. Boylan, E. A. Shale. 2003. An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. Journal of the Operational Research Society, 54 (8), 833–837.

Juran, D. C., L. W. Schruben. 2004. Using worker personality and demographic information to improve system performance prediction. Journal of Operations Management, 22 (4), 355–367.

Kahn, K. B. 1998. Revisiting top-down versus bottom-up forecasting. The Journal of Business Forecasting Methods & Systems, 17 (2), 14–19.

Kerr, N. L., R. S. Tindale. 2011. Group-based forecasting? A social psychological analysis. International Journal of Forecasting, 27 (1), 14–40.

Kirchgässner, G., U. K. Müller. 2006. Are forecasters reluctant to revise their predictions? Some German evidence. Journal of Forecasting, 25 (6), 401–413.

Kohn, R. 1982. When is an aggregate of a time series efficiently forecast by its past? Journal of Econometrics, 18 (3), 337–349.

Kremer, M., B. Moritz, E. Siemsen. 2011. Demand forecasting behavior: system neglect and change detection. Management Science, 57 (10), 1827–1843.

Kuo, Y., K. Liang. 2004. Human judgments in New York state sales and use tax forecasting. Journal of Forecasting, 23 (4), 297–314.

Kurz-Kim, J. 2008. Combining forecasts using optimal combination weight and generalized autoregression. Journal of Forecasting, 27 (5), 419–432.

Lau, N., S. Hasija, J. N. Bearden. 2014. Newsvendor pull-to-center reconsidered. Decision Support Systems, (58), 68–73.

Lawrence, M., P. Goodwin, M. O'Connor, D. Önkal. 2006. Judgmental forecasting: a review of progress over the last 25 years. International Journal of Forecasting, 22 (3), 493–518.

Lawrence, M., M. O'Connor. 1992. Exploring judgemental forecasting. International Journal of Forecasting, 8 (1), 15–26.

Lawrence, M., M. O'Connor. 2005. Judgmental forecasting in the presence of loss functions. International Journal of Forecasting, 21 (1), 3–14.

Lawrence, M., M. O'Connor, B. Edmundson. 2000. A field study of sales forecasting accuracy and processes. European Journal of Operational Research, 122 (1), 151–160.

Lee, H. L., V. Padmanabhan, S. Whang. 1997. The bullwhip effect in supply chains. Sloan Management Review, 38 (3), 93–102.

Lee, H. L., K. C. So, C. S. Tang. 2000. The value of information sharing in a two-level supply chain. Management Science, 46 (5), 626–643.

Levn, E., A. Segerstedt. 2004. Inventory control with a modified Croston procedure and Erlang distribution. International Journal of Production Economics, 90 (3), 361–367.

Li, R., J. K. Ryan. 2011. A Bayesian inventory model using real-time condition monitoring information: a Bayesian inventory model. Production and Operations Management, 20 (5), 754–771.

Lin, J., M. Vlachos, E. Keogh, D. Gunopulos. 2004. Iterative incremental clustering of time series. Advances in Database Technology-EDBT 2004. Springer, 106–122.

Lipscomb, J., G. Parmigiani, V. Hasselblad. 1998. Combining expert judgment by hierarchical modeling: an application to physician staffing. Management Science, 44 (2), 149–161.

Lütkepohl, H. 1984. Linear transformations of vector ARMA processes. Journal of Econometrics, 26 (3), 283–293.

Makridakis, S., R. L. Winkler. 1983. Averages of forecasts: some empirical results. Management Science, 29 (9), 987–996.

Massey, C., G. Wu. 2005. Detecting regime shifts: the causes of under-and overreaction. Management Science, 51 (6), 932–947.

McCarthy, T. M., D. F. Davis, S. L. Golicic, J. T. Mentzer. 2006. The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. Journal of Forecasting, 25 (5), 303–324.

McCarthy Byrne, T. M., M. A. Moon, J. T. Mentzer. 2011. Motivating the industrial sales force in the sales forecasting process. Industrial Marketing Management, 40 (1), 128–138.

Min, C., A. Zellner. 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics, 56 (1), 89–118.

Moon, M. A., J. T. Mentzer, C. D. Smith. 2003. Conducting a sales forecasting audit. International Journal of Forecasting, 19 (1), 5–25.

Moritz, B., E. Siemsen, M. Kremer. 2014. Judgmental forecasting: cognitive reflection and decision speed. Production and Operations Management, 23 (7), 1146–1160.

Moritz, B. B., A. V. Hill, K. L. Donohue. 2013. Individual differences in the newsvendor problem: behavior and cognitive reflection. Journal of Operations Management, 31 (1), 72–85.

Morris, P. A. 1977. Combining expert judgments: a Bayesian approach. Management Science, 23 (7), 679–693.

Nauta, A., C. K. W. De Dreu, T. van der Vaart. 2002. Social value orientation, organizational goal concerns and interdepartmental problem-solving behavior. Journal of Organizational Behavior, 23 (2), 199–213.

Nauta, A., J. de Vries, J. Wijngaard. 2001. Power and biased perceptions of interdepartmental negotiation behavior. Group Processes & Intergroup Relations, 4 (3), 263–270.

Nauta, A., K. Sanders. 2001. Causes and consequences of perceived goal differences between departments within manufacturing organizations. Journal of Occupational and Organizational Psychology, 74 (3), 321–342.

Newbold, P., C. W. J. Granger. 1974. Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society. Series A (General), 137 (2), 131–165.

O'Connor, M., W. Remus, K. Griggs. 2000. Does updating judgmental forecasts improve forecast accuracy? International Journal of Forecasting, 16 (1), 101–109.

Ogden, T. 1997. Essential wavelets for statistical applications and data analysis. Springer Science & Business Media.

Oliva, R., N. Watson. 2009. Managing functional biases in organizational forecasts: a case study of consensus forecasting in supply chain planning. Production and Operations Management, 18 (2), 138–151.

Oliva, R., N. Watson. 2011. Cross-functional alignment in supply chain planning: a case study of sales and operations planning. Journal of Operations Management, 29 (5), 434–448.

Önkal, D., M. Lawrence, K. Zeynep Sayım. 2011. Influence of differentiated roles on group forecasting accuracy. International Journal of Forecasting, 27 (1), 50–68.

Önkal, D., K. Zeynep Sayım, M. Lawrence. 2012. Wisdom of group forecasts: does role-playing play a role? Omega, 40 (6), 693–702.

Orcutt, G. H., H. W. Watts, J. B. Edwards. 1968. Data aggregation and information loss. The American Economic Review 773–787.

Palm, F. C., A. Zellner. 1992. To combine or not to combine? Issues of combining forecasts. Journal of Forecasting, 11 (8), 687–701.

Percival, D. B., A. T. Walden. 2006. Wavelet methods for time series analysis. Cambridge University Press.

Porras, E., R. Dekker. 2008. 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.

Proietti, T. 2000. Comparing seasonal components for structural time series models. International Journal of Forecasting, 16 (2), 247–260.

Protzner, S. 2015. Mind the gap between demand and supply: a behavioral perspective on demand forecasting. Ph.D. thesis, ERIM.

Raghunathan, S. 2001. Information sharing in a supply chain: a note on its value when demand is nonstationary. Management Science, 47 (4), 605–610.

Rapach, D. E., J. K. Strauss. 2008. Forecasting US employment growth using forecast combining methods. Journal of Forecasting, 27 (1), 75–93.

Ritzman, L. P., B. E. King. 1993. The relative significance of forecast errors in multistage manufacturing. Journal of Operations Management, 11 (1), 51–65.

Romeijnders, W., R. Teunter, W. van Jaarsveld. 2012. A two-step method for forecasting spare parts demand using information on component repairs. European Journal of Operational Research, 220 (2), 386–393.

Schultz, Carl R. 1987. Forecasting and inventory control for sporadic demand under periodic review. Journal of the Operational Research Society, 38 (5), 453–458.

Schwarzkopf, A. B., R. J. Tersine, J. S. Morris. 1988. Top-down versus bottom-up forecasting strategies. The International Journal of Production Research, 26 (11), 1833–1843.

Schweitzer, M. E., G. P. Cachon. 2000. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. Management Science, 46 (3), 404–420.

Shale, E. A., J. E. Boylan, F. R. Johnston. 2006. Forecasting for intermittent demand: the estimation of an unbiased average. Journal of the Operational Research Society, 57 (5), 588–592.

Singhal, J., K. Singhal. 2007. Holt, Modigliani, Muth, and Simon's work and its role in the renaissance and evolution of operations management. Journal of Operations Management, 25 (2), 300–309.

Småros, J. 2007. Forecasting collaboration in the European grocery sector: observations from a case study. Journal of Operations Management, 25 (3), 702–716.

Snyder, R. 2002. Forecasting sales of slow and fast moving inventories. European Journal of Operational Research, 140 (3), 684–699.

Snyder, R. D., J. K. Ord, A. Beaumont. 2012. Forecasting the intermittent demand for slow-moving inventories: a modelling approach. International Journal of Forecasting, 28 (2), 485–496.

Sohn, S. Y., M. Lim. 2007. Hierarchical forecasting based on AR-GARCH model in a coherent structure. European Journal of Operational Research, 176 (2), 1033–1040.

Solomou, S., M. Weale. 1991. Balanced estimates of UK GDP 1870-1913. Explorations in Economic History, 28 (1), 54–63.

Solomou, S., M. Weale. 1993. Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920-38. Journal of the Royal Statistical Society. Series A (Statistics in Society), 156 (1), 89.

Solomou, S., M. Weale. 1996. UK national income, 1920-1938: the implications of balanced estimates. The Economic History Review, 49 (1), 101–115.

Sterman, J. D. 1989. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. Management Science, 35 (3), 321–339.

Stone, R., D. G. Champernowne, J. E. Meade. 1942. The precision of national income estimates. The Review of Economic Studies, 9 (2), 111.

Struzik, Z. R. 2001. Wavelet methods in (financial) time-series processing. Physica A: Statistical Mechanics and its Applications, 296 (1), 307–319.

Su, X. 2008. Bounded rationality in newsvendor models. Manufacturing & Service Operations Management, 10 (4), 566–589.

Swanson, N. R., T. Zeng. 2001. Choosing among competing econometric forecasts: regression-based forecast combination using model selection. Journal of Forecasting, 20 (6), 425–440.

Syntetos, A. A., M. Z. Babai, N. Altay. 2012. On the demand distributions of spare parts. International Journal of Production Research, 50 (8), 1–17.

Syntetos, A. A., M. Z. Babai, Y. Dallery, R. Teunter. 2009a. Periodic control of intermittent demand items: theory and empirical analysis. Journal of the Operational Research Society, 60 (5), 611–618.

Syntetos, A. A., J. E. Boylan. 2001. On the bias of intermittent demand estimates. International Journal of Production Economics, 71 (1), 457–466.

Syntetos, A. A., J. E. Boylan. 2005. The accuracy of intermittent demand estimates. International Journal of Forecasting, 21 (2), 303–314.

Syntetos, A. A., J. E. Boylan. 2006. On the stock control performance of intermittent demand estimators. International Journal of Production Economics, 103 (1), 36–47.

Syntetos, A. A., J. E. Boylan, J. D. Croston. 2005. On the categorization of demand patterns. Journal of the Operational Research Society, 56 (5), 495–503.

Syntetos, A. A., N. C. Georgantzas, J. E. Boylan, B. C. Dangerfield. 2011. Judgement and supply chain dynamics. Journal of the Operational Research Society, 62 (6), 1138–1158.

Syntetos, A. A., K. Nikolopoulos, J. E. Boylan. 2010. Judging the judges through accuracy-implication metrics: the case of inventory forecasting. International Journal of Forecasting, 26 (1), 134–143.

Syntetos, A. A., K. Nikolopoulos, J. E. Boylan, R. Fildes, P. Goodwin. 2009b. The effects of integrating management judgement into intermittent demand forecasts. International Journal of Production Economics, 118 (1), 72–81.

Terwiesch, C., Z. J. Ren, T. H. Ho, M. A. Cohen. 2005. An empirical analysis of forecast sharing in the semiconductor equipment supply chain. Management Science, 51 (2), 208–220.

Teunter, R., B. Sani. 2009. On the bias of Croston's forecasting method. European Journal of Operational Research, 194 (1), 177–183.

Teunter, R. H., L. Duncan. 2009. Forecasting intermittent demand: a comparative study. Journal of the Operational Research Society, 60 (3), 321–329.

Teunter, R. H., A. A. Syntetos, M. Z. Babai. 2010. Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. European Journal of Operational Research, 203 (3), 619–624.

Teunter, R. H., A. A. Syntetos, M. Z. Babai. 2011. Intermittent demand: linking forecasting to inventory obsolescence. European Journal of Operational Research, 214 (3), 606–615.

Tiao, G. C., I. Guttman. 1980. Forecasting contemporal aggregates of multiple time series. Journal of Econometrics, 12 (2), 219–230.

Trapero, J. R., R. Fildes, A. Davydenko. 2011. Nonlinear identification of judgmental forecasts effects at SKU level. Journal of Forecasting, 30 (5), 490–508.

Walz, D. T., D. B. Walz. 1989. Combining forecasts: multiple regression versus a Bayesian approach. Decision Sciences, 20 (1), 77–89.

Wang, W., A. A. Syntetos. 2011. Spare parts demand: linking forecasting to equipment maintenance. Transportation Research Part E: Logistics and Transportation Review, 47 (6), 1194–1209.

Weale, M. 1985. Testing linear hypothesis on national account data. The Review of Economics and Statistics, 67 (4), 685.

Weale, M. 1988. The reconciliation of values, volumes and prices in the national accounts. Journal of the Royal Statistical Society. Series A (Statistics in Society), 151 (1), 211.

Wei, W. W. S., B. Abraham. 1981. Forecasting contemporal time series aggregates. Communications in Statistics-Theory and Methods, 10 (13), 1335–1344.

Wichard, J. D. 2011. Forecasting the NN5 time series with hybrid models. International Journal of Forecasting, 27 (3), 700–707.

Widiarta, H., S. Viswanathan, R. Piplani. 2007. On the effectiveness of top-down strategy for forecasting autoregressive demands. Naval Research Logistics, 54 (2), 176–188.

Widiarta, H., S. Viswanathan, R. Piplani. 2009. Forecasting aggregate demand: an analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. International Journal of Production Economics, 118 (1), 87–94.

Wiehenbrauk, D. 2010. Collaborative promotions, *Lecture Notes in Economics and Mathematical Systems*, vol. 643. Springer.

Willemain, T. R., C. N. Smart, 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. Shockor, P. A. DeSautels. 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. International Journal of Forecasting, 10 (4), 529–538.

Winkler, R. L., R. T. Clemen. 1992. Sensitivity of weights in combining forecasts. Operations Research, 40 (3), 609–614.

Winkler, R. L., S. Makridakis. 1983. The combination of forecasts. Journal of the Royal Statistical Society 150–157.

Wright, G., G. Rowe. 2011. Group-based judgmental forecasting: an integration of extant knowledge and the development of priorities for a new research agenda. International Journal of Forecasting, 27 (1), 1–13.

Yaniv, I. 2011. Group diversity and decision quality: amplification and attenuation of the framing effect. International Journal of Forecasting, 27 (1), 41–49.

Zellner, A. 1969. On the aggregation problem: a new approach to a troublesome problem. K. A. Fox, J. K. Sengupta, G. V. L. Narasimham, eds., Economic Models, Estimation and Risk Programming. Springer, 365–374.

Zhang, X. 2004. Evolution of ARMA demand in supply chains. Manufacturing & Service Operations Management, 6 (2), 195–198.

Zhao, X., J. Xie. 2002. Forecasting errors and the value of information sharing in a supply chain. International Journal of Production Research, 40 (2), 311–335.

Zhao, X., J. Xie, J. Leung. 2002. The impact of forecasting model selection on the value of information sharing in a supply chain. European Journal of Operational Research, 142 (2), 321–344.

Nederlandse Samenvatting

Voorspellingen zijn essentieel als gevolg van een inherent onzekere consumentenvraag. Retailers gebruiken voorspellingen voor verkoop-, voorraad- en inkoopbeslissingen; leveranciers voor productie- en aankoopbeslissingen; en distributeurs voor toewijzing van capaciteit. Voorspellingsfouten zijn aanzienlijk in de praktijk en hebben een negatieve impact op de operationele prestaties. Bedrijven missen de kennis, expertise en training op het gebied van voorspellen om hun besluitvorming goed te ondersteunen. Het verminderen of minimaliseren van voorspellingsfouten staat in dit proefschrift dan ook centraal. Dit wordt bereikt door het verbeteren van de voorspellingsmogelijkheden van bedrijven via zowel een uitbreiding van de beschikbare voorspellingsmethoden en modellen als een analyse van op welke manier het voorspellingsproces, de context waarin deze methoden en modellen worden ingezet, kan worden verbeterd.

Uitbreiding van de beschikbare voorspellingsmethoden en modellen wordt gedaan in twee verschillende studies, waarin reeds beschikbare informatie, zonder context-specifieke veronderstellingen, wordt benut om aanzienlijke winst te behalen. De eerste studie vult bestaande voorspellingsmethoden aan voor het type vraag waarin aanzienlijke tussenpozen voorkomen tussen vraagmomenten. Bestaande methoden negeren een mogelijke relatie tussen de verstreken tijd en de hoogte van toekomstige orders of richten zich op de risico's van incourante voorraden. Deze beperking heeft gevolgen voor de prestaties van het voorraadbeheer en het financiële resultaat. Dit proefschrift presenteert een nieuwe voorspellingsmethode die op basis van de verstreken tijd de toekomstige vraag kan anticiperen. Toepassingen op data over vijf verschillende probleemgebieden tonen aan dat dit aanzienlijke financiële voordelen biedt. De tweede studie generaliseert bestaande voorspellingsmethoden en vervangt de traditionele tegenstelling van top-down en bottom-up methoden door te onderzoeken hoe de hiërarchie van producten en categorieën kan worden gebruikt bij het voorspellen. Producten vormen van nature groepen in een hiërarchie met verkopen van individuele producten, productgroepen en categorieën, en de totale verkoop.

Twee veelgebruikte benaderingen beginnen vanaf tegenoverliggende einden van de hiërarchie, onder- en bovenaf, om voorspellingen te genereren voor al deze niveaus. Beide benaderingen impliceren verlies van informatie door het maken van afzonderlijke voorspellingen, die worden beschouwd als onafhankelijk. Door rechtstreeks gezamenlijke voorspellingen te genereren voor een groep producten wordt informatie beter benut. Dit vertaalt zich in superieure logistieke en financiële prestaties. Dit proefschrift presenteert een dergelijke aanpak die nadrukkelijk de mogelijkheid van productafhankelijkheden, zoals complementariteit van producten en productsubstitutie, behelst die anders worden genegeerd. Een simulatiestudie en een empirische studie tonen de aanzienlijke winst die hiermee behaald kan worden. Deze aanpak kan toegepast worden voor hiërarchische voorspellingen in het algemeen, en reikt verder dan de huidige toepassing. Waar de eerste studie beschikbare informatie benut voor afzonderlijke producten, doet de tweede studie dit door expliciet mee te nemen dat producten in groepen en hierarchieën vallen.

De praktische toepassing van voorspellingsmethoden vormt het onderwerp van de laatste twee studies, die het voorspellingsproces bij bedrijven analyseren en met name de rol van intuïtie. Beslissingen op basis van intuïtie vormen de kern van het voorspellingsproces bij veel bedrijven, en dit heeft een directe invloed op de prestaties. Deze studies geven inzicht in hoe het gedrag van voorspellers verschilt en hoe dit wordt beïnvloed. Dit proefschrift laat zien dat het gedrag systematisch verschilt in zoverre dat voorspellers in twee duidelijk verschillende soorten groepen kunnen worden ingedeeld, waarbij een groep te drastisch reageert op voorspellingsfouten en een andere te zwak reageert. De studies modelleren de voorspellingsbeslissingen die participanten tijdens uitgebreide experimenten maken op een manier die brede heterogeneiteit van participanten toestaat. Dit proefschrift onderzoekt ook de invloed van functies en financiële prikkels, en concludeert dat deze invloed aanzienlijk is, en ook onbewust het voorspellingsgedrag beïnvloedt. Ook test dit proefschrift het effect van een populair wegingsmechanisme om de voorspellingen van meerdere voorspellers te gebruiken en concludeert dat deze niet genoeg is om het effect van het ongewenste gedrag weg te halen. Een wegingsmechanisme dat rekening houdt met bewuste en onbewuste fouten lijkt betere prestaties te realiseren.

Het onderzoek in dit proefschrift verhoogt de voorspellingscapaciteiten van bedrijven door de beschikbare voorspellingsmethoden en modellen uit te breiden en door te laten zien hoe het voorspellingsproces, waarin deze methoden en modellen zijn verankerd, kan worden verbeterd. Doordat voorspellingen alle beslissingen ondersteunen zijn de financiële gevolgen groot.

Curriculum Vitae

Clint Pennings (1986) has a bachelor's and master's degree from Utrecht University, and a master's degree from Erasmus University Rotterdam. In 2011, Clint started his PhD research at the Rotterdam School of Management as part of the Dutch Institute for Advanced Logistics (Dinalog) project *4C4More*, a collaboration between several universities and companies. Within this project, he focused on extending available methods to generate more accurate forecasts and on modeling the behavior of forecasters to analyze how the forecasting process can be improved. He is passionate about working with companies to apply statistics and machine learning in their business processes.

Clint presented his research at various international academic conferences, such as EURO, IFORS, and INFORMS, and at several more practice-oriented venues. Next to research he greatly enjoys teaching in the bachelor and master programmes. Currently, he works as a postdoctoral researcher at the Rotterdam School of Management.

Author Portfolio

Articles

C.L.P. Pennings, J. van Dalen & E. van der Laan. Exploiting Elapsed Time for Managing Intermittent Demand for Spare Parts. Accepted for publication in *European Journal of Operational Research*.

C.L.P. Pennings & J. van Dalen. Collaborative Forecasting in (2015) *Cross-Chain Collaboration in the Fast Moving Consumer Goods Supply Chain*.

Under review

C.L.P. Pennings & J. van Dalen. Integrated Hierarchical Forecasting.

C.L.P. Pennings, J. van Dalen & L. Rook. Chasers, Smoothers and Departmental Biases: Heterogeneity in Judgmental Forecasting.

P. Bouman, **C.L.P. Pennings**, J. van Dalen & L. Kroon. Time Choice Data for Public Transport Optimization.

Working papers

C.L.P. Pennings, J. van Dalen & L. Rook. Coordinating Judgmental Forecasting: Coping with Intentional Biases.

C.L.P. Pennings. Competitive Markets Online: Pricing Policies and Price Dispersion.

C.L.P. Pennings. Personalized Product Maps for Large Webshops.

Teaching

MSc thesis supervision	*Supply Chain Master*
Big Data and Business Analytics	*Business Information Management Master*

Decision Support	*Bachelor*
Forecasting in the Supply Chain	*Supply Chain Master*
Research Methods	*Master*
Statistical Methods	*Bachelor*

PhD Courses

Advanced Econometrics III
Applied Bayesian Statistics
Bayesian Econometrics
Cooperative Games
Data Analysis and Statistics
Discrete Choice Modelling
Econometric Analysis
Markov Decision Processes
Multi-Agent Systems Research
Noncooperative Games
Panel Data Econometrics
Probabilistic Modelling
Publishing Strategy
Statistical Methods
Stochastic Models and Optimisation

Conference Presentations

2012	**EURO Conference**	Vilnius, Lithuania
2012	**LIS Workshop**	Brussels, Belgium
2013	**EURO Conference**	Rome, Italy
2013	**OR Conference**	Rotterdam, the Netherlands
2013	**INFORMS Conference**	Minneapolis, Minnesota, USA
2013	**LIS Research Summit**	Rotterdam, the Netherlands
2013	**TRAIL Conference**	Delft, the Netherlands
2014	**IFORS Conference**	Barcelona, Spain
2014	**Dinalog Conference**	Amsterdam, the Netherlands
2015	**EURO Conference**	Glasgow, Scotland
2015	**CASPT Conference**	Rotterdam, the Netherlands
2015	**4C4More Conference**	Amersfoort, the Netherlands

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

Dissertations in the last five years

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotors: Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor: Prof. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: strategy, structure and performance*, Promotors: Prof. J.Y.F. Wynstra & Dr E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor: Prof. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotors: Prof. H.G. Barkema & Prof. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotors: Prof. U. Kaymak & Prof. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor: Prof. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotors: Prof. H.W. Volberda & Prof. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor: Prof. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670

Benschop, N, *Biases in Project Escalation: Names, frames & construal levels*, Promotors: Prof. K.I.M. Rhode, Prof. H.R. Commandeur, Prof. M.Keil & Dr A.L.P. Nuijten, EPS-2015-375-S&E, hdl.handle.net/1765/79408

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor: Prof. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor: Prof. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Beusichem, H.C. van, *Firms and Financial Markets: Empirical Studies on the Informational Value of Dividends, Governance and Financial Reporting*, Promotors: Prof. A. de Jong & Dr. G. Westerhuis, EPS-2016-378-F&A, http://repub.eur.nl/pub/93079

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor: Prof. J. Veenman & Prof. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotors: Prof. A.G.Z. Kemna & Prof. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivationsand Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotors: Prof. H.G. Barkema & Dr D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotors: Prof. G.A. van der Knaap & Prof. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur.nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit,Coworker Satisfaction, and Relational Models*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision-Making*, Promotor: Prof. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotors: Prof. F. Langerak & Prof.G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotors: Prof. H.A.M. Daniels & Prof. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor: Prof. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Cranenburgh, K.C. van, *Money or Ethics: Multinational corporations and religious organisations operating in an era of corporate responsibility*, Prof. L.C.P.M. Meijs, Prof. R.J.M. van Tulder & Dr D. Arenas, EPS-2016-385-ORG, http://repub.eur.nl/pub/93104

Consiglio, I., *Others: Essays on Interpersonal and Consumer Behavior*, Promotor: Prof. S.M.J. van Osselaer, EPS-2016-366-MKT, http://repub.eur.nl/pub/79820

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor: Prof. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Darnihamedani, P. *Individual Characteristics, Contextual Factors and Entrepreneurial Behavior*, Promotors: A.R. Thurik & S.J.A. Hessels, EPS-2016-360-S&E, http://repub.eur.nl/pub/93280

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor: Prof. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor: Prof. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotors: Prof. J.J.P. Jansen, Prof. F.A.J. van den Bosch, & Prof. H.W. Volberda, EPS-2012-258-STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?* Promotor: Prof. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor: Prof. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duyvesteyn, J.G. *Empirical Studies on Sovereign Fixed Income Markets,* Promotors: Prof. P.Verwijmeren & Prof. M.P.E. Martens, EPS-2015-361-F&A, hdl.handle.net/1765/79033

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor: Prof. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor: Prof. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Elemes, A, *Studies on Determinants and Consequences of Financial Reporting Quality,* Promotor: Prof. E.Peek, EPS-2015-354-F&A, http://hdl.handle.net/1765/79037

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor: Prof. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Erlemann, C., *Gender and Leadership Aspiration: The Impact of the Organizational Environment*, Promotor: Prof. D.L. van Knippenberg, EPS-2016-376-ORG, http://repub.eur.nl/pub/79409

Eskenazi, P.I., *The Accountable Animal*, Promotor: Prof. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotors: Prof. J. van Oosterhout & Prof. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction under Prominence*, Promotor: Prof. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Faber, N., *Structuring Warehouse Management*, Promotors: Prof. MB.M. de Koster & Prof. Ale Smidts, EPS-2015-336-LIS, http://repub.eur.nl/pub/78603

Fernald, K., *The Waves of Biotechnological Innovation in Medicine: Interfirm Cooperation Effects and a Venture Capital Perspective,* Promotors: Prof. E.Claassen, Prof. H.P.G. Pennings & Prof. H.R. Commandeur, EPS-2015-371-S&E, http://hdl.handle.net/1765/79120

Fliers, P.T., *Essays on Financing and Performance: The role of firms, banks and board*, Promotor: Prof. A. de Jong & Prof P.G.J. Roosenboom, EPS-2016-388-F&A, http://repub.eur.nl/pub/93019

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotors: Prof. J.J.P. Jansen & Prof. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gaast, J.P. van der, *Stochastic Models for Order Picking Systems*, Promotors: Prof. M.B.M de Koster & Prof. I.J.B.F. Adan, EPS-2016-398-LIS, http://repub.eur.nl/pub/93222

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor: Prof. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor: Prof. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor: Prof. B.G.C. Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotors: Prof. G.M.H. Mertens & Prof. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hekimoglu, M., *Spare Parts Management of Aging Capital Products,* Promotor: Prof. R. Dekker, EPS-2015-368-LIS, http://hdl.handle.net/1765/79092

Heij, C.V., *Innovating beyond Technology. Studies on how management innovation, co-creation and business model innovation contribute to firm's (innovation) performance*, Promotors: Prof. F.A.J. van den Bosch & Prof. H.W. Volberda, EPS-2012-370-STR, http://repub.eur.nl/pub/78651

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor: Prof. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotors: Prof. F.A.J. van den Bosch & Prof. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor: Prof. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, A.C., *Sentiment Analysis of Text Guided by Semantics and Structure,* Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2015-369-LIS, http://hdl.handle.net/1765/79034

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotors: Prof. U. Kaymak & Prof. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hollen, R.M.A., *Exploratory Studies into Strategies to Enhance Innovation-Driven International Competitiveness in a Port Context: Toward Ambidextrous Ports,* Promotors: Prof. F.A.J. Van Den Bosch & Prof. H.W.Volberda, EPS-2015-372-S&E, hdl.handle.net/1765/78881

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotors: Prof. H.P.G. Pennings & Prof. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotors: Prof. D. de Cremer & Dr M. van Dijke, EPS-2011-244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotors: Prof. P.J.F. Groenen & Prof. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotors: Prof. D. de Cremer & Prof. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotors: Prof. L.G. Kroon & Prof. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: A Neuroscientific Approach*, Promotor: Prof. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor: Prof. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor: Prof. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor: Prof. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor: Prof. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotors: Prof. G.A. van der Knaap & Prof. H.P.G. Pennings, EPS-2011-223-ORG, http://repub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor: Prof. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor: Prof. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotors: Prof. F.M. Go & Prof. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotors: Prof. H.W.G.M. van Heck & Prof. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotors: Prof. G.M.H. Mertens & Prof. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotors: Prof. J. Paauwe & Dr L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotors: Prof. S.L. van de Velde & Prof. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor: Prof. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor: Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotors: Prof. B. Wierenga & Prof. S.M.J. van Osselaer, EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Lee, C.I.S.G, *Big Data in Management Research: Exploring New Avenues*, Promotors: Prof. S.J. Magala & Dr W.A. Felps, EPS-2016-365-ORG, http://repub.eur.nl/pub/79818

Legault-Tremblay, P.O., Corporate Governance During Market Transition: Heterogeneous responses to Institution Tensions in China, Promotor: Prof. B. Krug, EPS-2015-362-ORG, http://repub.eur.nl/pub/78649

Lenoir, A.S. *Are You Talking to Me? Addressing Consumers in a Globalised World,* Promotors: Prof. S. Puntoni & Prof. S.M.J. van Osselaer, EPS-2015-363-MKT, http://hdl.handle.net/1765/79036

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotors: Prof. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor: Prof. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof. H. Bleichrodt, Prof. P.P. Wakker, & Prof. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor: Prof. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotors: Prof. H.R. Commandeur & Dr K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotors: Prof. A.R. Thurik, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotors: Prof. J. Spronk & Prof. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Ma, Y., *The Use of Advanced Transportation Monitoring Data for Official Statistics,* Promotors: Prof. L.G. Kroon and Dr Jan van Dalen, EPS-2016-391-LIS, hdl.handle.net/1765/80174

Manders, B., *Implementation and Impact of ISO 9001*, Promotor: Prof. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor: Prof. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Mell, J.N., *Connecting Minds: On The Role of Metaknowledge in Knowledge Coordination*, Promotor: Prof. D.L. van Knippenberg, EPS-2015-359-ORG, http://hdl.handle.net/1765/78951

Meuer, J., *Configurations of Inter-firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor: Prof. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor: Prof. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotors: Prof. J.J.P. Jansen, Prof. F.A.J. van den Bosch, & Prof. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor: Prof. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Moniz, A, *Textual Analysis of Intangible Information*, Promotor(s): Prof C.B.M. van Riel, Prof. F.M.G de Jong & Dr G.A.J.M. Berens, EPS-2016-393-ORG, http://repub.eur.nl/pub/93001

Mulder, J. *Network design and robust scheduling in liner shipping*, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, Eps-2016-384-LIS, http://repub.eur.nl/pub/80258

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Neerijnen, P., *The Adaptive Organization: the socio-cognitive antecedents of ambidexterity and individual exploration*, Promotors: Prof. J.J.P. Jansen, P.P.M.A.R. Heugens & Dr T.J.M. Mom, EPS-2016-358-S&E, http://repub.eur.nl/pub/93274

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in short term planning and in disruption management*, Promotor: Prof. L.G. Kroon, EPS-2011-224-LIS, http://repub.eur.nl/pub/22444

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotors: Prof. G.J. van der Pijl, Prof. H.R. Commandeur & Prof. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Oord, J.A. van, *Essays on Momentum Strategies in Finance,* Promotor: Prof. H.K. van Dijk, EPS-2016-380-F&A, hdl.handle.net/1765/80036

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on bureaucracy and formal rules*, Promotor: Prof. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Ozdemir, M.N., *Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances*, Promotor: Prof. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor: Prof. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotors: Prof. P.J.F. Groenen & Prof. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor: Prof. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How organizations negotiate permanent uncertainty and unresolvable choices*, Promotors: Prof. P.P.M.A.R. Heugens & Prof. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/pub/26392

Pourakbar, M., *End-of-Life Inventory Decisions of Service Parts*, Promotor: Prof. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotors: Prof. H.J.H.M. Claassen & Prof. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Protzner, S. *Mind the gap between demand and supply: A behavioral perspective on demand forecasting,* Promotors: Prof. S.L. van de Velde & Dr L. Rook, EPS-2015-364-LIS, *http*://repub.eur.nl/pub/79355

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor: Prof. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietdijk, W.J.R. *The Use of Cognitive Factors for Explaining Entrepreneurship,* Promotors: Prof. A.R. Thurik & Prof. I.H.A. Franken, EPS-2015-356-S&E, http://repub.eur.nl/pub/79817

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger, Prof. P.J.F. Groenen, & Prof. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotors: Prof. A.G.Z. Kemna & Prof. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Rösch, D. *Market Efficiency and Liquidity*, Promotor: Prof. M.A. van Dijk, EPS-2015-353-F&A, http://hdl.handle.net/1765/79121

Roza, L., *Employee Engagement In Corporate Social Responsibility: A collection of essays*, Promotor: L.C.P.M. Meijs, EPS-2016-396-ORG, http://repub.eur.nl/pub/93254

Roza-van Vuren, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of innovation, absorptive capacity and firm size*, Promotors: Prof. H.W. Volberda & Prof. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor: Prof. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schoonees, P. *Methods for Modelling Response Styles*, Promotor: Prof.dr P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/79327

Schouten, M.E., *The Ups and Downs of Hierarchy: the causes and consequences of hierarchy struggles and positional loss*, Promotors; Prof. D.L. van Knippenberg & Dr L.L. Greer, EPS-2016-386-ORG, http://repub.eur.nl/pub/80059

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor: Prof. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Smit, J. *Unlocking Business Model Innovation: A look through the keyhole at the inner workings of Business Model Innovation*, Promotor: H.G. Barkema, EPS-2016-399-S&E, http://repub.eur.nl/pub/93211

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotors: Prof. D.L. van Knippenberg & Dr D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor: Prof. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor: Prof. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor: Prof. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor: Prof. D.L. van Knippenberg & Prof. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor: Prof. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor: Prof. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision-Making*, Promotor: Prof. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor: Prof. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotors: Prof. F.A.J. van den Bosch, Prof. H.W. Volberda, & Prof. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor: Prof. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor: Prof. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Valogianni, K. *Sustainable Electric Vehicle Management using Coordinated Machine Learning*, Promotors: Prof. H.W.G.M. van Heck & Prof. W. Ketter, Eps-2016-387-LIS, http://repub.eur.nl/pub/93018

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor: Prof. L.G. Kroon, EPS-2014-327-LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Vermeer, W., *Propagation in Networks:The impact of information processing at the actor level on system-wide propagation dynamics,* Promotor: Prof. P.H.M.Vervest, EPS-2015-373-LIS, http://repub.eur.nl/pub/79325

Versluis, I., *Prevention of the Portion Size Effect*, Promotors: Prof. Ph.H.B.F. Franses & Dr E.K. Papies, EPS-2016-382-MKT, http://repub.eur.nl/pub/79880

Vishwanathan, P., *Governing for Stakeholders: How Organizations May Create or Destroy Value for their Stakeholders*, Promotors: Prof. J. van Oosterhout & Prof. L.C.P. M. Meijs, EPS-2016-377-ORG, http://repub.eur.nl/pub/93016

Visser, V.A., *Leader Affect and Leadership Effectiveness: How leader affective displays influence follower outcomes*, Promotor: Prof. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor: Prof. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Vries, J. de, *Behavioral Operations in Logistics,* Promotors: Prof. M.B.M de Koster & Prof. D.A. Stam, EPS-2015-374-LIS, http://repub.eur.nl/pub/79705

Wagenaar, J.C., *Practice Oriented Algorithmic Disruption Management in Passenger Railways*, Prof. L.G. Kroon & Prof. A.P.M. Wagelmans, EPS-2016-390-LIS, http://repub.eur.nl/pub/93177

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotors: Prof. R. Dekker & Prof. U. Kaymak, EPS-2011-248-LIS, http://repub.eur.nl/pub/26564

Wang, P., *Innovations, status, and networks*, Promotors: Prof. J.J.P. Jansen & Dr V.J.A. van de Vrande, EPS-2016-381-S&E, http://repub.eur.nl/pub/93176

Wang, T., *Essays in Banking and Corporate Finance*, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor: Prof. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor: Prof. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotors: Prof. H.R. Commandeur & Prof. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor: Prof. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor: Prof. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Yuferova, D. *Price Discovery, Liquidity Provision, and Low-Latency Trading*, Promotors: Prof. M.A. van Dijk & Dr D.G.J. Bongaerts, EPS-2016-379-F&A, http://repub.eur.nl/pub/93017

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor: Prof. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor: Prof. I. Dittmann, EPS-2012-261-F&A, http://repub.eur.nl/pub/32344

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotors: Prof. A.R. Thurik & Prof. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422

The demand that drives various activities in the supply chain is inherently uncertain, necessitating the need for forecasting. Retailers require forecasts for sales, inventory and order decisions, suppliers for production and procurement decisions, and distributors for capacity allocation decisions. In practice, forecast errors are substantial, which negatively affects operational performance. Reducing or minimizing these forecast errors is central to this thesis and is achieved by improving the forecasting capabilities of companies, which encompasses extending the available forecasting methods as well as analyzing how the forecasting process, the context in which these methods and models are embedded, can be improved.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl