Modelling and Managing Supply Chain Forecast Uncertainty in the Presence of the Bullwhip Effect

Patrick Saoud

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the Department of Management Science, Lancaster University

December 2019

Marketing Analytics & Forecasting | Lancaster University Management School

Declaration

This thesis is my own work and it has not been submitted in support of an application for another higher degree or qualification elsewhere.

Patrick Saoud

Abstract

The Bullwhip Effect, defined as the upstream amplification of demand variability, has received considerable interest in the field of Supply Chain Management in recent years. This phenomenon has been detected in various industries and sectors, and manifests itself with multiple inefficiencies and higher costs at upper echelons in the supply chain. As a result, this topic is of great importance for academics and practitioners alike. One root cause of the Bullwhip Effect is the need for firms to forecast demand in order to place their orders and base their inventory decisions. Despite the multitude of studies that have emerged tackling this issue, the impact of the quality of forecasts on the Bullwhip Effect has received limited coverage in the literature. Modelling and forecasting the demand can be challenging, resulting in increased forecast uncertainty that contributes to the Bullwhip Effect.

This thesis aims at bridging this gap by investigating three main research questions: (i) How can supply chain forecast uncertainty be captured at a firm level? (ii) How can the upstream propagation of forecast uncertainty from the Bullwhip Effect be measured? and (iii) What customer demand information sharing strategy is the most effective in reducing upstream the forecast uncertainty and inventory costs resulting from the Bullwhip Effect?

We first propose an empirical approximation for measuring forecast uncertainty at a local level, which we show to outperform commonly used approximations for inventory purposes. We then propose a novel metric to capture the propagation of forecast uncertainty at higher echelons in the Supply Chain, which correlates strongly with upstream inventory costs, more so than the conventional Bullwhip measure. Using this, we evaluate alternative information sharing strategies that have appeared in the literature, but have not been assessed comparatively. We find that relying solely on point of sales data results in the best forecasting accuracy and inventory cost performance for upstream members. The findings obtained are actionable and simple to implement, making them of great use and relevance for supply chain practitioners and managers.

Acknowledgments

Pursuing a PhD can seem at times a daunting and onerous task. Opting for an academic path means that one has to often sit idly drowning in a sea of papers while observing the world evolve at its never-ending pace. But as my supervisor Nikos has incessantly and rightly reminded me over the years, it is the ideal time to truly reflect and contemplate on one's life. As a student nearing his thirties on the verge of graduating, I can honestly say that this adventure has allowed me to count my blessings, and to be grateful for the people whose presence has graced my life, as they each have independently been an integral source of inspiration in my quest for personal growth.

I would first like to thank my parents Hani and Christiane for always providing me with their unconditional support throughout this experience. I sincerely have been blessed to have them, as they always managed to outdo themselves as exceptional parents, friends and human beings. My sister Samar has time and time again stood besides me, mechanically checking up on me on a daily basis, and always bending backwards to find the sneakiest and most humorous ways to help me take the edge off my PhD. I can't thank her enough for giving me in her husband Joe Rahme the older brother I never had, and for making me a godfather to their adorable daughter Zoe. I would also like to express my gratitude towards my Uncle Ziad and his wife Danielle for being second parents to me, as well as my godfather Tony Tadros, as I can't begin to imagine a life without them in it.

I have no doubt in my mind that I have been extremely fortunate in having my supervisors, Nikolaos Kourentzes and John E. Boylan, oversee my work during this time. Never one to shy away from a challenge, Nikos has always pushed me to tackle problems nobody had cared or dared to address, to seek their practical implications and relevance to real-life situations, and to always look at the bigger picture. With more frequent flyer miles than the most overworked cabin pilot (despite his life-long relationship with the Mediterranean Sea), he has surprisingly always been able to afford me and his other students his unfettered attention, time and comments. But apart from the invaluable academic lessons he so staunchly inculcated in me, I ended up learning other worthy life-related principles from him as well:

that complacency equates with mediocrity, that no situation is devoid of a comical aspect, that humans can survive without oxygen but not without ice cream, and that maintaining a balance in any task one undertakes is crucial for self-improvement. As I pen these words, I can't claim to adhere to the latter two just yet, but I have my whole life ahead of me. John has been a remarkable supervisor, juggling between being an accomplished academic with a wide network of collaborators, a supervisor to many students, as well as head of the Management Science programme. He has always managed to find ways to improve any research I embarked on with his breadth of insights and information, and has always encouraged me to reach greater heights in my work. A true gentleman with an unmatched sense of etiquette and tactfulness, as well as an impeccable grasp on the English language and its grammar, I fear that my time at Lancaster was not nearly enough to learn all that I could from him. I am immensely grateful to the both of them for their patience, support and efforts, and for turning me into a better researcher and person by upholding the highest academic standards and work ethics.

This experience would not have been possible without the scholarship granted by the Management School, for which I really am grateful. As a member of the Centre for Marketing Analytics and Forecasting, I was lucky to be part of the team, being surrounded by supportive and helpful colleagues. However, the people who made the biggest impression on me had to be the staff members that I encountered along my journey. What can be said about Robert Fildes that hasn't been said already? He has received a lot of praise over the years from many academics and students alike for his clairvoyance and enormous wealth of knowledge (as well as his ability to knock down any research project with his trademark "so what?" question), but his modesty and sense of humour are the qualities that have struck me the most as memorable. Adam Hindle has always been a true inspiration: dynamic, innovative, affable and excellent at conveying any message, he truly is a gem for the Department, and I am certain that the majority of his previous and current students share my opinion of him. Nicos Pavlidis is without a doubt an exceptional academic, capable of tackling any problem, all while being a well-rounded and really agreeable character to grab a beer with, and discuss a wide array of subjects. Ivan Svetunkov (and of course his lovely wife Anna Sroginis), with his joviality, his signature fedora hat, and his constant passion for the field, is a reminder to anyone that following your dreams is key to leading a happy life (if in doubt, ask him what his cat ended up being named). I would also like to take this chance to thank the marvellous trio of Gay Bentinck, Jackie Clifton and Lindsay Newby for the immeasurable assistance they have constantly offered me throughout my academic path. They truly are the backbone of the Management Science department at Lancaster University, and I am sure that expressing my gratitude towards all their efforts merely scratches the surface in explaining their contribution to me or any other student during their studies.

This list of acknowledgments wouldn't be complete without a tribute to my wonderful friends, whom I unequivocally consider family. Raed Daou, with his 16 pillars of "Parara" wisdom, has continuously offered me support throughout the mental ebbs and flows of the PhD, always striving to take my mind off academia, and lifting my morale up in times when even I deemed it to be an impossible task. I owe him a huge debt of gratitude, and the term brother is an understatement of the bond we share. In the end, who else other than his delightful brother Maher would have opened my eyes to the untapped world of dolphin-shaving? His parents Sami and Mervat have always embraced me as their third son, and I hold them with the highest esteem. Ali Zeiour has to be the warmest and most genuine person I have met, always pushing me to be true to myself, and to keep my head high no matter what. Lyn is very lucky to have him as a father. Rakel Lárusdóttir and her two adorable sons (Stefan and my best friend Óliver) have always reminded me that there is no obstacle one can't surmount without a smile and positive attitude. I have the uttermost respect and affection for her parents Stefan and Guðrun, for the candour and congeniality they have always displayed towards me during my numerous stays in sunny Iceland. Alina and Polina Ilinova have undoubtedly had an everlasting impact on me, as I am forever indebted to them both for moulding me into the man I am today. And finally, to Farah Bahmad, life can really be cruel sometimes, having snatched him away from everyone so early. He was a real angel, constantly smiling and laughing, and not a day goes by where I don't think of him and of how proud he would have been of my achievements. For what it's worth, this PhD is dedicated to him, may he rest in peace.

Lancaster, December 2019

Patrick

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation and Context

With the increase in market competitiveness, maintaining good supply chain operations is key for businesses to deliver the end-products to customer. To achieve this, the flow of information and material between the different supply chain members must be managed efficiently. One problem that many supply chains experience is the Bullwhip Effect, defined as the upstream amplification of demand variability (Lee et al., 1997b). This poses a challenge to the different members of the supply chain, especially at higher tiers, as the upstream members perceive a more erratic demand than at the customer level.

The amplification in demand variability is associated with many deleterious consequences, as it generates production swings, higher inventory and transportation costs, and increases in lost sales and customer dissatisfaction (Lee et al., 1997b; Towill et al., 2007; Haughton, 2009). Several case studies have documented the existence of the Bullwhip Effect (for e.g., Hammond, 1994; Holweg et al., 2005; Terwiesch et al., 2005), and empirical studies have detected its presence in numerous industrial sectors (for e.g., Akkermans and Voss, 2013; Isaksson and Seifert, 2016; Jin et al., 2017). Given the wide extent of its consequences, as well as its prevalence in practice, this topic has sparked considerable interest from researchers in the field of supply chain management, and is of key importance for practitioners aiming to meet their performance targets and improving their operations at a local and global level of the supply chain.

In the literature, many factors have been determined to cause the Bullwhip Effect (Bhattacharya and Bandyopadhyay, 2011). One of them is demand signal processing, which refers to the need for firms to forecast future demand in order to produce their ordering and inventory decisions (Lee et al., 1997b). At the downstream level, retailers who observe customer demand must forecast in order to plan their orders. These orders are passed to subsequent echelons in the supply chain, serving as the demand for the upstream members, and they have been found to be more variable than the initial demand, which in turn results in the Bullwhip Effect. Hence, the quality of forecasts impacts the upstream magnification of demand variability, since it determines the volatility of the upstream incoming demand signals, and thus is crucial in understanding the Bullwhip Effect (Fildes et al., 2008).

There exists some challenges associated with forecasting demand in a supply chain context. For inventory purposes, forecasts are produced over a planning horizon in order to cover lead-time demand and to set safety stocks. Despite its importance, this topic has not received an adequate coverage in the literature (Syntetos et al., 2016). One of the reasons firms hold inventories is to buffer against the uncertainty surrounding their future demand, and this is in the form of safety stocks. Higher levels of demand and forecast uncertainty result in additional inventory being held (Rumyantsev and Netessine, 2007), which impacts negatively the inventory turnover and the firm's profitability (Gaur et al., 2005a; Hançerlioğulları et al., 2016). Poor forecasts result in inadequate safety stocks level for a firm, which incur higher inventory costs and lower service levels (Liao and Chang, 2010; Sanders and Graman, 2009; Kerkkänen et al., 2009), as well as more volatile orders (Zhang, 2004a). Better forecasts can allow managers to mitigate the upstream amplification of orders (and thus the Bullwhip Effect), as well as reduce unnecessary inventory costs.

Despite the empirically-verified relationship between sales and inventory levels (for e.g., Granger and Lee, 1989; Kesavan et al., 2010), there are gaps in the current understanding of the interaction between forecasting and stock control (Syntetos et al., 2009). This thesis aims at exploring further the relationship between sales and inventories, in order to obtain a better grasp of its link to the Bullwhip-related inventory costs in a supply chain. Even though these associations have been established at an empirical level, we still do not fully understand the complex interaction between forecasting and inventory control. In addition, many of the studies dedicated to the Bullwhip Effect have relied on restrictive assumptions, which limit the applicability of their results (Miragliotta, 2006). One assumption encountered in many papers is full knowledge of the underlying customer demand process. This however does not shed light on the impact of mis-specifying forecasts, which is common in standard practice (Fildes and Kingsman, 2011). Thus, the impact of forecast uncertainty on the stock control performance will be examined in this thesis, in order to offer practical and actionable insights.

One suggested method to reduce the upstream amplification of demand variability is the upstream sharing of customer demand information (Lee et al., 2000). Since members at upper echelons do not observe the original demand, but instead incoming orders from their immediate downstream partners, this exchange of information is expected to mitigate the impact of the Bullwhip, as the downstream demand signal, unaffected by the information distortions resulting from the Bullwhip, should allow the upstream members to get a clearer view of the original customer demand. In practice, sharing information is costly for supply chain partners, and several barriers are associated with its implementation (Kembro and Näslund, 2014; Kembro et al., 2014; Kembro and Selviaridis, 2015). In the literature, opposing views have emerged with regards to the value of information sharing; however a few studies have quantitatively validated its benefits (Kemppainen and Vepsäläinen, 2003; Småros, 2007; Fildes et al., 2008). The benefits associated with information sharing should thus be carefully examined before any agreement takes place, as it is of paramount importance for practitioners seeking to engage in such collaboration schemes. Even though many studies have dealt with this topic, a limited number of researchers have compared alternatives to the theoretical approach often used in the upstream forecasting process, which substitutes incoming orders from their downstream partner with the original customer demand signal for the orders received, and none of these studies has addressed the inventory implications of different strategies for information sharing.

1.2 Research Questions

Throughout this thesis, we investigate the case of a Make-To-Stock supply chain with fast moving items. The following research questions are formulated:

- *Research Question 1: How to measure and capture lead time forecast uncertainty at* *the firm level?* In an inventory setting, forecasts are produced over a lead time, and this results in their errors being correlated with each other over that interval. This correlation accrues in the estimation of the lead time variance of forecast error, which is necessary for safety stock estimation. An adequate measure for forecasting uncertainty, which accounts for these correlations, must be established to determine appropriate safety stock levels, and to get a better understanding on the impact of the former on the inventory performance of a firm.

- *Research Question 2: How to measure the upstream propagation of forecast uncertainty resulting from the Bullwhip Effect?* The quality of forecasts has been determined to be a contributor to the presence of the Bullwhip Effect in supply chain. As downstream forecast uncertainty impacts upstream members by distorting the original customer demand signal, studying how forecast errors are transmitted along the supply chain allows us to explore its relationship with the demand variability amplification from the Bullwhip Effect and its related inventory costs.

- *Research Question 3: What is the impact of different modelling strategies on reducing forecast uncertainty and inventory costs?* Downstream demand information sharing has been identified as a measure to counter the amplification of demand variability, as it allows upstream members to observe the customer demand signal undistorted by the factors which contribute to the Bullwhip Effect. In light of the new measures proposed in the previous research questions, the effectiveness of different information sharing strategies on upstream forecasting accuracy with respect to their incoming demand signal will be re-evaluated in the context of a decentralised supply chain (where each entity behaves to minimise their own costs), as well as its impact on upstream inventory costs, in order to assess which strategy for using customer demand information is the most effective, and whether indeed information sharing is beneficial for upstream members.

1.3 Contributions

1.3.1 Chapter 2

To hedge against demand uncertainty, firms carry additional inventory in the form of safety stocks. These are determined by calculating the variance of forecast errors over lead time, and since various demand processes require different estimates of the latter, approximations are often used in practice. However, some of these are theoretically inadequate, as they ignore the correlation of forecast errors that accumulates over lead time. Chapter 2 reviews different approximations for the estimation of lead time forecast errors variance, explaining their theoretical underpinnings and highlighting analytically their drawbacks. It then proceeds to propose a new empirical approximation, and compares its inventory performance to the current ones under different demand uncertainty settings. Earlier versions of this chapter have been presented at the *EURO 2015*, *International Society for Inventory Research 2016* and *EURO 2017* conferences. A manuscript of this chapter is available as a working paper (Saoud et al., 2018), and has been submitted to the *European Journal of Operational Research* and is currently under review.

1.3.2 Chapter 3

A key component in studying the Bullwhip Effect is its measurement, in order to determine whether the phenomenon is present, and whether a proposed solution is effective in taming it. The currently adopted measure links back to its definition and consists of the ratio of upstream to downstream demand variances. However, this measure suffers from a few drawbacks which are often encountered in practice. In addition, demand variability is only a measure of spread of the data, while it is the uncertainty related to predicting demand that is the cost driver. Chapter 3 thus delineates the difference between both concepts of demand uncertainty and variability, that have been used interchangeably in the literature. It then suggests a new metric, based on the approximation proposed in Chapter 2, that measures the upstream propagation of forecast uncertainty, an established cause of the Bullwhip Effect. It then compares the relationship of the two with upstream inventory costs. Earlier versions of this chapter have been presented at the *International Society for Inventory Research 2018*. A manuscript of this chapter is available as a working paper (Saoud et al., 2019), and has been submitted to *International Journal of Production Economics*.

1.3.3 Chapter 4

Information sharing has been advocated as a potential method to alleviate the negative consequences of the Bullwhip Effect, as it allows upstream members to gain visibility over the downstream demand information. In the literature, the findings have pointed in opposite directions regarding its benefits. Furthermore, a limited number of studies have compared different approaches for exchanging information. Chapter 4 thus reviews and compares different strategies for sharing information, and reassesses its value on forecasting accuracy by employing the metric derived in Chapter 3, and on inventory costs, in the presence of forecast uncertainty and managerial adjustments made to final ordering decisions. This is currently a working paper in preparation for submission.

1.4 Research Methodology and Modeling Approaches

In this section, the research methodology adopted throughout this thesis is discussed. First, we cover the different modeling approaches featured, highlighting their respective strengths and weaknesses, as well as identifying how each method is employed in our research. Next, the model verification and validation schemes are presented, in order to assess the validity of the deployed models and their derived findings.

From the research questions posed in the previous sections, a quantitative research methodology is required as the overarching goal in this thesis lies in measuring, modeling and managing forecast uncertainty, as well as quantifying its inventory impact. The concept of uncertainty is pervasive in many problems in the sciences (Briggs, 2016), and there is no universally accepted method to measure it (Jurado et al., 2015). In addition, conflicting views have emerged in the literature on how to model forecast uncertainty (Fildes, 1985; Hendry et al., 1990; Mingers, 2006; Chiasson et al., 2006). We therefore aim at presenting a consistent modeling methodology to address this issue.

Within the field of Management Science, there exists different research methodology paradigms, each underpinned by its philosophical implications, and each having garnered various criticisms as a result (Ackoff, 1979; van Gigch, 1989; Churchman, 1994; Meredith, 2001; Mingers, 2003). With the multitude of methodologies that can be adopted, and the drawbacks associated with each, relying exclusively on any single one can hinder the validity of the results obtained in our research. Hence, more than one modeling approach is required to overcome some of the limitations of any specific methodology. In this thesis, a hybrid quantitative methodology comprised of three different approaches is used, which consist of: (i) the analytical approach, (ii) the simulation approach, and (iii) the empirical approach.

1.4.1 Analytical Approach

Also referred to as axiomatic research, this strand of research involves building a conceptual model to represent the problem at stake and usually relies on mathematical analysis to reach its conclusions (Bertrand and Fransoo, 2002). By imposing certain conditions, the researcher is able to study the relationship between different variables of choice, while isolating the effect of others and can derive closed-form solutions or theoretical bounds for the studied problem. The obtained results hold as long as the premised assumptions are true, but their validity and usefulness is contingent on the model being representative of the real-life situation (Breiman, 2001). Numerous influential papers in the Bullwhip Effect literature have followed this path, shedding light on several triggers and potential remedies to this phenomenon (e.g. Lee et al., 1997b; Dejonckheere et al., 2003).

Despite its frequent use, there exists some weaknesses that are associated with this methodological approach. For instance, as the size and degree of complexity of the problem increases, its solution becomes analytically intractable. As a result, other methods are required to in order to address the research questions. But apart from this concern, a critical aspect with this type of research lies in the strength of the assumptions made. Indeed, the assumptions can be restrictive, and may thus not adequately reflect the real life situation. Hence, many theoretical results might not be implemented in practice, which can be attested for example in the disparity between theoretical inventory models and those employed in practice (Silver, 1981; Cattani et al., 2011). Fildes and Kingsman (2011) warn that many researchers on the Bullwhip Effect have opted to trade off pragmatic problems for those offering elegant solutions and mathematical tractibility, thus harming the managerial relevance of their results. In other cases, the theory governing the studied system might not be fully understood, and as a result, the researchers might be unaware that they are imposing assumptions on the model which do not hold practically.

In this thesis, the analytical approach is used mainly for theoretical exposition purposes. Indeed, it is employed in Chapter 2 to highlight the presence of non-zero covariance terms between forecast errors over the lead time even when the demand process and its parameters are assumed to be known, as well as to quantify their contribution to the lead time conditional variance of the forecast errors. In Chapter 3, we resort to this methodology when discussing the concept of forecast uncertainty and the different components that it is comprised of. Given that our research is concerned with forecasting uncertainty, there exists multiple ways to mis-specify the forecasting model. In order to avoid being restricted to representing this uncertainty with a pre-specified incorrect model, the analytical approach is supplemented further by the methodologies that follow.

1.4.2 Simulation Approach

As mentioned previously, one of the drawbacks encountered with the analytical approach relates to the complexity of the problem. Computer simulation is an effective method to tackle this issue, being more flexible than its counterpart as it enables larger models to be examined (Pidd, 2009). Owing to the complex and dynamic nature of supply chains, it has proven to be a useful tool for modeling supply chain problems (Van Der Zee and Van Der Vorst, 2005). By employing this method, researchers attempt to mimic the real life situation and its inherent randomness, and reaches their conclusions through statistical analysis of the outputs (Law, 2014). It allows them to exert full control over the design of the model and the variables included, thus easily overcoming the tractability problem faced by the analytical approach (Harrison et al., 2007). When used in conjunction with the analytical approach, it can serve the purpose of corroborating the findings derived in the latter.

While this method is advocated as an effective way to augment the analytical approach for complex situations, it nonetheless shares the same criticism as it counterparts, namely that the results obtained depend on how representative the latter is of the real world situation (Flynn et al., 1990). The findings obtained via simulation are not a general proof, and are thus conditional on the model's assumptions and mechanism (Bertrand and Fransoo, 2002). In some cases, the researcher might be unaware of any additional assumptions that may be implied. In addition, simulation models should undergo a thorough examination to assess the soundness of the model and thus verify and validate it (this is elaborated further in Section 2.4).

In this thesis, the simulation methodology is adopted as the main modeling approach, given the benefits it offers over the analytical one. It is suitable to model forecast uncertainty in a generic way, as different mis-specified models can be fit to represent the underlying demand. In addition, it enables full flexibility and control over the design of the supply chain model, and enables the study of more advanced demand processes, as well as more variables that might contribute to either forecasting uncertainty or the Bullwhip Effect in general.

1.4.3 Empirical Approach

Both the analytical and simulation approach discussed so far are theoretical methodologies, which raises the question of how grounded in reality are the findings from either methodologies. A third stream of research methodology exists, the empirical approach, where the researcher relies on real life data rather than synthetic one to reach their conclusions (Flynn et al., 1990). There are different empirical approaches to research. In this thesis it is discussed within the context of a single firm in the form of empirical simulation (Shafer and Smunt, 2004). Under this type of study, which is typically more complex than the previous two, the assumptions imposed previously are relaxed, and the researcher can test if their understanding of the studied problem holds when these assumptions are violated, either weakly or strongly (Bertrand and Fransoo, 2002). Thus it can be used as a means to assess the validity of theoretical findings, by deploying those with real data, as well as their robustness to deviations from the ideal situation.

There are some drawbacks to employing an empirical approach. For instance, gathering the data can be costly, and it might not always be readily available (Flynn et al., 1990). In addition, the findings derived pertain to the firm under study and are not general. From a methodological perspective, this approach does not have the control over the model design features that was found in the previous two. The lack of transparency in the real demand or operational system may lead to a limited understanding of the findings when they do not follow theory. Furthermore, there will be many confounding factors within the model, due to the complexity of the real system.

In this thesis, the empirical approach is used in a supporting role to validate theoretical findings. It features in Chapter 2, where it is employed alongside both the analytical and simulation approach to confirm the results obtained in that chapter. As stated above, the biggest obstacle to adopting this approach is the acquisition of adequate data for the research questions at hand.

1.5 Model Verification and Validation

The main modeling approach adopted throughout this thesis relies on a supply chain simulation, as argued for in Section 1.4. The usefulness and verisimilitude of a simulation model largely depends on it being backed by theoretical foundations and empirical evidence, and this poses a challenge to researchers as there exists no universal set of criteria and approaches to guarantee this (Naylor and Finger, 1967). This is especially true since a model is not universal but designed for specific purposes, which implies that establishing its credibility will vary according to the purpose itself (Robinson, 2014). Nonetheless, some procedures have been adopted to assess the soundness and accuracy of a simulation, mainly through model verification and validation (Pidd, 2009). Model verification is the process of ensuring the correct implementation of the conceptual model, while model validation assesses the accuracy of the model in representing the real system being studied. Both concepts and their application in this thesis are elaborated in this section.

1.5.1 Model Verification

Model verification is concerned with ensuring that the developed model accurately represents the conceptual one (Pidd, 2009). It typically involves checking that the computer model is devoid of any programming errors, and that its logical structure is sound. In this thesis, this was accomplished in several ways. Good programming practice and rigorous debugging were applied to the different components and subcomponents that constitute the model at each step of its development. Trace variables, where a list of detailed variables, counters and calculations is recorded after each event occurs in the simulation, were also employed to guarantee that the model was running as planned (Law, 2014). In addition, intermediate outputs were estimated manually at different steps of the simulation and compared with the output produced from the latter. Furthermore, the simulation was conducted under simplified cases where analytical solutions exist in order to further verify the soundness of the model (Kleijnen, 1995). Finally, the pseudo-random number generators were tested via visual inspection as well as statistical tests to confirm the stochastic behaviour of the studied system (Sargent, 2013).

1.5.2 Model Validation

The purpose of model validation is to determine whether the simulation adequately represents the studied system. In this thesis, three approaches were implemented: (i) conceptual model validation, (ii) white-box and black box validation, and (iii) experimentation validation. These are discussed below.

1.5.2.1 Conceptual Model Validation

The purpose of conceptual validation is to establish that the conceptual model underlying the simulation possesses a sound logical and theoretical foundation, and that its assumptions are reasonable and adequate to represent the problem at stake (Sargent, 2013). It typically involves feedback from other experts to jointly assess the conceptual model (Robinson, 2014), but this is not possible in this case. Instead, the conceptual model's structure and assumptions have built on earlier published work to verify that they were suitable for the research questions posed in this thesis. Whilst these model assumptions are not unreasonable, the possibility of mis-specification has been taken into account, as discussed earlier, making them a more reasonable set of conceptual models.

1.5.2.2 White-Box and Black-Box Validation

Under black-box validation, the modeller assumes the model's mechanisms are unknown, and compares by means of statistical tests the output generated by the simulation with real data collected from the reference system, or with output from an alternative model (Pidd, 2009). This type of validation is performed after the final model has been developed, and it focuses on the predictive power of the simulation, as it assesses how accurately the final output resembles its real life counterpart (Sargent, 2013; Robinson, 2014). White-box validation on the other hand assumes full visibility of the internal mechanisms of the model, and aims at establishing that each of the model's contents represent accurately enough to those from the real system (Pidd, 2009). It bears similarities with model verification, as the latter is concerned with checking that the developed simulation model follows the conceptual model, and as a result both share many of the same procedures, such as the use of event traces and the comparison between simulation outputs and known analytical solutions (Robinson, 2014).

In our research, black-box validation was not feasible, given the complexity of the simulated items and the unavailability of real data and alternative models. White-box validation was conducted using the techniques described in Section 1.5.1. One additional issue that was addressed was the validation of the generated time series that serve as demand inputs in the models. There exists no guaranteed set of methods to ascertain whether a generated series adheres to a specific demand process, and so time series plots, such as the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were first inspected to check if the series exhibited the theoretical properties prescribed for those processes. However, these rely on the asymptotic properties of these demand processes, and thus if a plotted series fails to display them, we are unable to discern whether this is caused by statistical sampling or by the incorrect generation of the data. Consequently, we resort to grey-box validation, where partial knowledge of the functioning of the model is assumed (Holst et al., 1993). This is achieved by using information criteria such as Akaike's Information Criterion (AIC, Akaike, 1974) to determine whether the generated time series follows indeed the process from which it was generated. Both methods concluded that in the majority of cases, the input time series were valid in representing their underlying theoretical processes. As the remainder of the cases were generated by the same process, we attribute any deviations to sampling uncertainty.

1.5.2.3 Experimentation Validation

This type of validation is concerned with setting up the appropriate experimental design procedures for the simulation in order to obtain reliable results (Robinson, 2014). Our simulation is non-terminating, as there exists no closing event in the system under study for which the experiment is stopped. As a result, we are interested in the simulation converging to its steady-state behaviour (if such exists) before any analysis is performed on its output. This is achieved via two ways: (i) conducting the simulation under a large number of replications, and (ii) allowing the simulation to run on a warm-up period (Law, 2014; Robinson, 2014).

Having the model run over a large number of replications is a crucial element for any simulation, as the results from a single or few replications are not sufficient to draw any inferences about the model. Indeed, the results from one replication are akin to the realisations of a single random variable from a sample with some variance, and might thus be different from the mean behaviour of the distribution from which it was drawn (Law, 2014). Therefore, many replications are necessary in order to dampen the variation between each run, in order for the Law of Large numbers to take effect and for the mean of the replications to converge to the population mean of the studied output.

In addition to the use of a large sample size and multiple replications, a warm-up period is also employed to ensure the experimental validation (Kleijnen, 1995). As certain subcomponents of the model have to be initialised (such as the inventory policy), a bias is incurred which can affect the collected results from the simulation. Moreover, the studied outputs require a certain number of iterations to occur before reaching their normal behaviour. Therefore, the simulation is allowed to first run on a burn-in or warm-up set, and this set of observations is subsequently truncated. In this thesis, two types of warm-up sets were employed. The first serves the purpose of eliminating the initialisation bias due to the generation of the demand time series. The second is necessary for the calculation of the conditional variance of forecast errors to be stable, as well as to remove the bias from initialising the safety stocks in the inventory policy. Furthermore, this burn-in set allows the studied outputs from the inventory policy to converge to their normal behaviour, such as estimates for the service levels or inventory costs. This set is inserted after the training set, where the forecasting model and parameters are estimated, and before the test set, where the model's output is collected and analysed. There exists no exact method to ascertain the number of observations necessary for this warm-up set (Schruben et al., 1983; Robinson, 2014), and therefore this was determined by means of visual inspection of the different outputs over multiple replications.

To better illustrate the experimentation validation procedures and their usefulness, consider the following simple example of calculating the cycle service level for an inventory simulation at the retailer level. The studied output is for a single replication in the model, and tracks the cumulative calculation of the average service level across observations, plotted in Figure 1.1. For this example, the downstream demand is generated as a first-order autoregressive AR(1) time series given by $y_t = 100 + 0.7y_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim N(0,1)$. An Order-Up-To inventory policy is used, with a review window of one period, and instantaneous order replenishments. The target cycle service level is a 90% coverage rate, and the initial safety stock is set at 200 units, which is twice the level of the demand. For illustrative purposes, 150 observations are produced, which are split equally into a training, burn-in and test set of 50 each.

In the training set, we first observe a decrease in the mean service level as the initial impact of the safety stock initialisation is fading off. However, the training set is insufficient on its own to obtain a steady estimate, as the computed service level fluctuates due to the small number of observations used in its calculation. Therefore, it requires an additional set of points in order for the estimate to stabilise itself. The simulation is thus allowed to run for further observations in the burn-in set before converging to its true value at the end of the test set. This example also highlights the importance of setting large sample sizes for the simulation, as more observations will enable the service level estimate to better converge to its population mean, as well as the importance of running a large number of replications, since different runs will produce different service level estimates which will vary from the one plotted in the above graph.

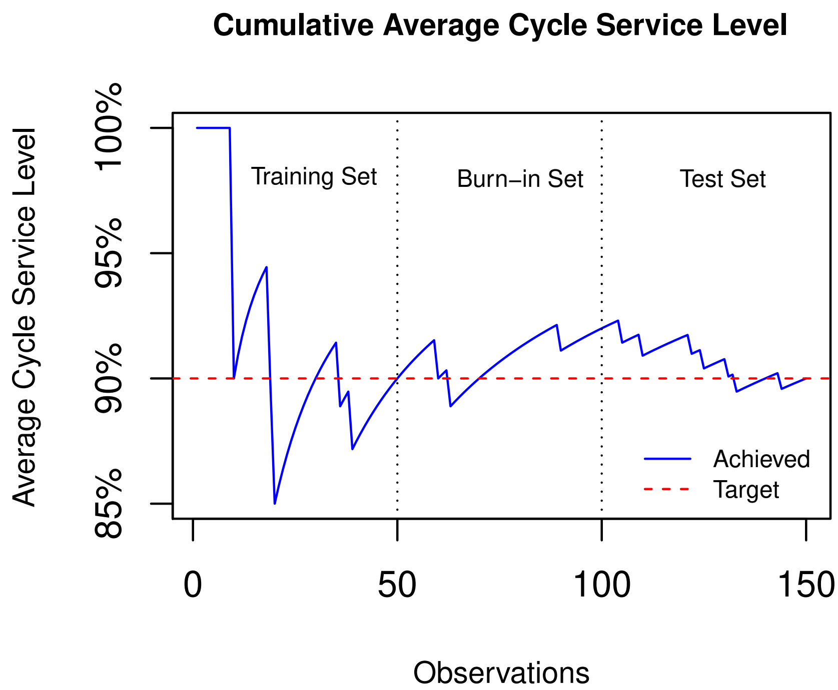

Figure 1.1: Example of the cumulative average cycle service level across observations for a single simulation replication.

1.6 Outline

The remainder of this thesis is organised as follows; Chapter 2 proposes an empirical approximation to capture demand uncertainty for safety stock estimation purposes. Drawing on this approximation, a novel metric is proposed to measure the upstream propagation of forecast uncertainty due to the Bullwhip in Chapter 3. Chapter 4 utilises the measure suggested in the previous chapter to contrast and evaluate the impact of different information sharing strategies. Finally, Chapter 5 discusses the contributions to the literature and their managerial implications, then offers venues for future research to expand on the findings from this thesis.

Chapter 2

Approximations for the Lead Time Variance: a Forecasting and Inventory Evaluation

Safety stock is necessary for firms in order to manage the uncertainty of demand. A key component in its determination is the estimation of the variance of the forecast error over lead time. Given the multitude of demand processes that lack analytical expressions of the variance of forecast error, an approximation is needed. It is common to resort to finding the one-step ahead forecast errors variance and scaling it by the lead time. However, this approximation is flawed for many processes as it overlooks the autocorrelations that arise between forecasts made at different lead times. This research addresses the issue of these correlations first by demonstrating their existence for some fundamental demand processes, and second by showing through an inventory simulation the inadequacy of the approximation. We propose to monitor the empirical variance of the lead time errors, instead of estimating the point forecast error variance and extending it over the lead time interval. The simulation findings indicate that this approach provides superior results to other approximations in terms of cycle-service level. Given its lack of assumptions and computational simplicity, it can be easily implemented in any software, making it appealing to both practitioners and academics.

2.1 Introduction

With an increase in competitiveness, meeting customer demand has become a target that businesses strive to achieve. Maintaining a balance between excess inventory and lost sales is a necessity for better performance. Safety stock plays a pivotal role in this, as it allows buffering against demand uncertainty. In practice, safety stock is determined by multiplying the standard deviation of forecast error over lead time by the inverse of the distribution function that represents these errors at the desired level of coverage. The forecast error over lead time implies knowledge of the cumulative forecast over the same period for which variance expressions are not readily available. This leads researchers and practitioners to either impose severe assumptions or use approximations to arrive at the desired variance by using the point forecast errors. Textbooks often prescribe scaling the variance for the one-step ahead forecast errors by the lead time as an approximation (Axsäter, 2015). Due to its simplicity and ease of application, this formula is employed frequently, requiring only the calculation of the one-step-ahead forecast errors variance as an input. This method suffers from a serious drawback as it fails to capture the correlations between forecast errors (Johnston and Harrison, 1986; Barrow and Kourentzes, 2016). Not accounting for these correlations leads to the variance of errors being under-estimated, which in turn results in inappropriate safety stocks levels being determined.

This research acknowledges the existence of these correlations and provides some new insights on their effect on forecasts, service levels and inventory holdings. While these correlations are the motivation for this chapter, the research aims at examining the performance of three different forecasting methods for the variance of lead time forecast error on stock control levels under various types of forecasting uncertainty. Second, within an AutoRegressive Integrated Moving Average (ARIMA) framework, we examine analytically simple fundamental Normal demand processes and use the variance-covariance matrix to illustrate the validity of the traditional approximation for an ARIMA(0,0,0) model but highlight the approximation's inadequacy for other commonly applied models due to the appearance of these correlations. This occurs even when the demand model and parameters are assumed to be fully known, as they appear in the variance-covariance matrix of the forecast errors, irrespective of the assumptions on the distribution of the latter, one of which is typically independence. The form of these correlations depends on the model structure and its parameters. Third, we also recommend the use of a simple yet intuitive heuristic for approximating the forecast errors variance, which consists of monitoring the forecast errors distribution over the lead time, rather than resorting to the approximation discussed before. While this approach has appeared before in the literature, the underlying motivation behind its choice has not been discussed, and this chapter seeks to justify its use and advantage over other approximations. The heuristic relies on the empirical cumulative forecast errors over lead time directly, instead of building on the one (or multiple) step ahead forecast errors. In contrast with existing research (e.g. Prak et al., 2017), we do not seek to determine remedies for specific demand processes where the forecasting model is mis-specified; rather we aim at studying the performance of approximations for estimating the lead time variance of forecast errors. Since knowledge of the underlying process is impossible in practice, we proceed to examine three competing approximations, explaining the rationale behind them and linking them to the analytical insights we provide. This chapter contributes to the existing literature by evaluating the inventory implications of these approximations, under different forecasting uncertainty settings and data generating processes, including both stationary and non-stationary ones, discussing the conditions under which each approximation is viable. We validate our analytical and simulation results on a real case study, using data from a US retailer. Our empirical findings indicate that this method achieves superior service level results compared to alternatives, which coupled with its ease of implementation, underlines its usefulness for research and practice.

The rest of this chapter is organised as follows: section 2.2 first discusses demand uncertainty and the approaches to model it; second, it covers the correlations between forecast errors over a lead time and how these impact the estimation of the variance of lead time forecast errors. Section 2.3 shows the existence of these correlations from a theoretical standpoint for simple yet fundamental demand processes. Section 2.4 examines the different approximations in estimating lead time forecast errors variance, while section 2.5 reports the findings drawn from an inventory simulation on the adequacy of these approximations, followed by the results gained from using real data in section 2.6.

2.2 Background Literature

2.2.1 Demand Uncertainty and Variability

Demand uncertainty refers to the unpredictability that arises in forecasting future demand, as opposed to variability which is defined as the fluctuations of demand around its mean. It is represented by the distribution of the forecast errors, and its impact on safety stocks is quantified by the variance of the latter. Failure to acknowledge this difference results in setting inappropriate safety stocks. The forecast uncertainty can be split into three types, which are reviewed subsequently. We show in the next section that these uncertainties appear in the calculation of the variance of lead time demand forecast errors.

In modeling demand, we are confronted with three types of uncertainty, endemic to any forecasting problem: *model, parameter* and *sample size* (Chatfield, 1995). Any forecasting task faces model uncertainty, as in reality the underlying Data Generating Process (DGP) is unknown and it is impossible to diagnose how closely this is approximated by a specified model. Many forecasters and inventory researchers overlook this and fail to account for this uncertainty, which is reflected in an inadequate estimation of the variance of forecast error. As a result, inappropriate safety stocks are set and higher inventory costs are incurred (Badinelli, 1990; Kim and Ryan, 2003; Dong and Lee, 2003). Even if the form of the true DGP is assumed to be known, it is questionable whether the parameters can be perfectly known. Parameter Uncertainty refers to this, where the parameters are misspecified, which yields an impact on the calculation of the variance via the existence of a bias of estimators (Ansley and Newbold, 1981), which can affect the performance of demand prediction intervals (Lee and Scholtes, 2014) as well as safety stocks (Ritchken and Sankar, 1984). Sample Size Uncertainty refers to the case where both model and parameters are known, and the uncertainty is from sampling issues, which results in a disparity between the asymptotic properties of the model and the finite sample properties of the underlying data. This can manifest itself in the behaviour of the error distribution (Phillips, 1979). In practice, it is often difficult to distinguish between the three types of uncertainty. For example, a misspecified parameter can make a model term insignificant, and change the specified model as well.

In a stock control context, quantiles are required in addition to point forecasts in order to model forecast uncertainty. Uncertainties, which generate biases, are not always unfavourable, as a parameter being optimal in terms of Mean Squared Error (MSE) is not optimal in an inventory context (Janssen et al., 2009), due to the difference in objective functions (Strijbosch et al., 2011). In fact, Silver and Rahnama (1986, 1987) and Janssen et al. (2009) adjust for the quantile estimation errors in order to achieve better service levels. As the discussed uncertainties appear in any forecasting task, they emerge in inventory problems where forecasts are required. However, for safety stock purposes, the variance of the forecast errors is required as an input. Since the error distribution is part of the forecasting model, the uncertainty surrounding the size of its standard deviation should be factored in, as it is estimated rather than determined a priori. Thus, all three uncertainties include this estimation uncertainty. While the impact of these uncertainties has featured separately in stock control papers, they have not been compared in terms of inventory performance. In this research, these uncertainties will be considered, in order to assess the effect of the safety stock approximations under these different uncertainty scenarios.

The uncertainty surrounding demand is quantified by constructing demand prediction intervals to confine the possible regions between which future demand might lie. The first step consists of determining the variance of the forecast errors. However, since many of the tasks consist of finite and limited data, the conditional variance of the i-th step ahead forecast error is employed ($\sigma^2_{t+i|t}$), as opposed to the unconditional one (σ^2_{t+i}) which provides the asymptotic value. Henceforth, the variance calculated is conditional on the data made available up until the estimation time t. This is the theoretical variance; nevertheless, when forecasting, the variance component of the forecast errors needs to be calculated, and thus $\hat{\sigma}^2_{t+1|t}$ replaces $\sigma^2_{t+1|t}$ for estimation purposes. After determining the variance component, the intervals are built with the use of the corresponding percentile of the assumed error distribution. The construction of demand or prediction intervals falls into three categories. The parametric stream of the research assumes that the underlying DGP can be modeled by a forecasting model (Lee, 2014), which depends on the researcher having adequately approximated the true model. For example, the Exponential Smoothing family of models can be used to manage SKUs with different patterns (Snyder et al., 2002), while knowing that it might not be optimal for all (if any) the time series; nevertheless its lead time expressions are derived for inventory purposes (Snyder et al., 2004). The non-parametric stream refrains from imposing any assumption on the demand process; and instead exploits the observed properties of the forecasting error density function. Examples consist of Chebyshev's Inequality (Gardner, 1988) or bootstrapping methods, with the latter finding successful use in approximating the intervals of a known demand process (Thombs and Schucany, 1990), and also performing well in setting reorder points (Wang and Rao, 1992) and meeting service levels (Fricker and Goodhart, 2000). The semi-parametric stream uses a mixture of the first two to model the intervals, such as those found in Taylor and Bunn (1999) and Lee and Scholtes (2014).

Demand prediction intervals have been studied more in the forecasting context than in inventory; however prediction intervals and safety stocks bear similarities, with the former referring to the area underneath a two-tailed interval of the error distribution, and the latter to that of a one-tailed statistical test. This relation between the two, while not identical, does allow the research on safety stocks to draw on that of prediction intervals. However, a major difference exists between the two: whereas prediction intervals rely on the variance of point forecast errors, safety stocks require the variance of errors over the lead time. This transition from point forecasts, which includes one period, to lead time forecasts, which comprises the several periods making up the lead time window, result in one of the difficulties faced in estimating the variance component for safety stocks, and this is elaborated in the subsequent section.

2.2.2 Safety Stock Estimation

The estimation of safety stocks requires the variance of forecast errors as input, as typically captured by the Mean Squared Error (MSE). Underlying this are two implicit assumptions. First, the forecasts are unbiased or have very little bias, and from the Bias-Variance decomposition of the MSE, this allows the variance of the errors to be approximated by the MSE (Wagner, 2002). This problem has been addressed by Lee and Scholtes (2014) for estimating better prediction intervals, and featured at the heart of the studies by Manary and Willems (2008) and Manary et al. (2009), who examined its impact on reorder points for Intel's inventories. The other assumption is that the errors are homoscedastic, which has been challenged as industrial data has exhibited some evidence of heteroscedasticity (Zhang, 2007; Stößlein et al., 2014; Trapero et al., 2019a,b).

There is another problem in estimating the forecast error variance over lead time. Indeed, forecasts are produced for several consecutive horizons in the future, akin to an overlapping temporal aggregation of demand itself over a window equal to the lead time (Boylan and Babai, 2016). This implies that the variance of the errors should cover the lead time window as well, and this brings forth a new issue which is the correlation of errors over lead times or horizons (up until this point, the terms "lead times" and "forecasting horizons" have been employed interchangeably. In this chapter, the term "horizon" is used to denote the protection interval for the safety stocks, and the exact relationship between these two in a periodic review context is Horizon = Lead Time + Review Period, while in a continuous review context, Horizon = Lead Time).

In the literature there are lead time forecast errors variance expressions for the well-known models; however with the multitude of possible demand processes, the majority of cases remain uncharacterised, and therefore a heuristic is required to circumvent this limitation. Standard textbooks recommend calculating the one-step-ahead forecast errors variance and multiplying it by the lead time, or in its more familiar form: $L\widehat{\sigma}^2_{t+1|t}$. This approximation depends only on the estimated variance of the one-step ahead forecast errors, which under perfect information of the demand pattern, reduces to the model innovations. No other parameters are required, such as the demand DGP autoregressive or moving average parameters. At first sight, its simplicity and lack of assumption of DGP makes its use appealing. Chatfield and Koehler (1991) criticised this method, with their main line of attack being that the concept of lead time forecasts is often confounded with that of forecasting for a horizon. Chatfield (1993) warns against the use of this approximation, stating that it possesses no theoretical basis and does not accommodate the different properties of the prediction intervals. Koehler (1990) challenges the robustness of this method, pointing out that this equation holds in the presence of the Random Walk model for multiple-steps-ahead point forecast errors. This approximation however is the cumulative L-steps-ahead conditional variance for an independent and identical demand (i.i.d) process, and it hinges on the crucial assumption that the demand being studied can be approximated in that fashion, thus ignoring other time series patterns such as autocorrelation, trend, moving averages or seasonal patterns. Furthermore, it is questionable whether the estimated MSE in-sample errors approximated well the out-of-sample variance of the forecast errors, an important distinction that is lost if the conditionality of errors is not considered (Barrow and Kourentzes, 2016).

An important distinction must be made between model errors (i.e. the process underlying the data) and forecast method errors, as these two terms are sometimes confused. The model errors or innovations at time t, ε_t are those found in the DGP; they form the stochastic component of the demand process. The forecasting or method residuals or errors at time t, e_t, measure the difference between actuals and forecasts. Under perfect knowledge of the demand, the forecast method errors are a function of the model structure, its errors and the set of parameters Θ, i.e. $e_{t+i|t} = f(\varepsilon_{t+1}, ...\varepsilon_{t+i}, \Theta)$, and this causes the correlations between multiple-steps-ahead errors to appear for several demand models (this is elaborated further in Section 2.3). Indeed, the main caveat of the current procedure for estimating variance is that it fails to capture two sources of correlations: (i) correlations between errors at different forecasting horizons and (ii) correlations between forecasts over cumulative lead time (Barrow and Kourentzes, 2016). The first refers to the correlations due to mis-specifying the forecasting model as i.i.d, while the second refers to those that arise due to the forecast errors at different lead times being correlated with each other. These correlations are the emphasis of the discussion in this chapter, as they are often overlooked in many of the estimation procedures, which ensues in inaccurate safety stocks levels.

For the first source of correlation, an implicit assumption being made is that the multiple-steps-ahead forecast errors all possess equal variance, with $\sigma^2_{t+i|t} = \sigma^2_{t+1|t}$. This assertion holds for the i.i.d process, $y_t = \mu + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$ and $\mathrm{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}) = 0$ for $i \neq j$. For other processes, $\sigma^2_{t+i|t} \neq \sigma^2_{t+1|t}$, as can be seen for the multiple-steps-ahead error equation for many DGPs. This already renders the approximation inappropriate. Kourentzes (2013) suggests a simple remedy to overcome this issue; namely to sum up the variance of forecast errors at different steps-ahead. Indeed, this approach consists of first estimating the conditional variance of the forecasting errors at each until the desired lead time L, and then adding them up to determine the variance at L, to give $\sum_{i=1}^{L} \sigma^2_{t+i|t}$. This approach is independent of any model assumptions, as it retrieves all the multiple-steps-ahead errors until the lead time and adds them up, and hence tackles the first issue posed by the $L\widehat{\sigma}^2_{t+1|t}$ approximation. This however does not address the correlation of forecasts at different lead times, as summing up the individual components implies automatically that the covariance terms between errors is 0.

The second source of correlation was acknowledged by Box et al. (2015) who proved its existence for ARIMA processes. The confusion between model innovations and forecast errors has partly led to these correlations being overlooked in the literature. Since the forecasts are produced for several steps-ahead and then summed, they are likely to be correlated as they share some of the information from previous periods, irrespective of whether the correct model has been identified or not. Johnston and Harrison (1986) used a Dynamic Linear Model formulation to show that ignoring these correlations could lead to an understatement of the variance of lead-time demand. They noted that this correlation existed, and they highlighted its omission from the ordinary methods of estimating the cumulative variance, attributing this correlation partly due to the need to estimate the level of the series, as well as the parameters. Prak et al. (2017) showed its existence under i.i.d demand, and provided correction terms for fitting both Simple Moving Average and Simple Exponential Smoothing procedures for this particular process, which delivered better safety stocks and service levels. While their correction terms focused on the i.i.d case and can not be extended to other demand processes, their work showed the impact of fitting a wrong model to a specific demand (i.e. the presence of correlations between errors due to model uncertainty). Nonetheless, even under perfect knowledge of the demand, and for different processes, the correlations may still exist, due to the model structure and parameters, and this is detailed further in Section 2.3.

2.3 Theoretical Derivations of the Variance and Covariance Terms

In this section, we examine the conditional variance of the lead time forecast errors analytically for certain fundamental demand processes and compare it to the typical $L\widehat{\sigma}^2_{t+1|t}$ approximation, with the objective of demonstrating the inadequacy of the latter and quantifying any ensuing losses from its use. This entails extracting the variance over lead time of these processes of the errors, as well as the covariance term between the errors, thus providing a unifying framework for the expressions of the demand processes studied. The covariance terms can then be used to form the Error Variance-Covariance matrix (Σ), which can display all these terms as well as demonstrate the impact of the accrual of these correlations and highlight how existing approximations capture or overlook these. We focus on DGPs that stem from the ARIMA(p,d,q) family of linear time series model, propounded by (Box et al., 2015). Its general expression is $\nabla^d y_t = \varepsilon_t \Theta(B)/\Phi(B) + c$, where y_t and ε_t represent the demand and the model innovations at time t, B the backshift operator, ∇, the difference operator defined as $\nabla = (1 - B)$, c the level or constant term, and $\Theta(B)$ and $\Phi(B)$ connote respectively the moving average and autoregressive operators in their polynomial form. The innovations ε_t adhere to an independent white noise process $N(0, \sigma^2)$. The ARIMA family encompasses many demand models studied in the literature and is common in the context of inventory control (Aviv, 2003), supporting our selection for this analysis.

While many textbooks have chapters covering the variance of point forecast errors for these processes, the cumulative aspect has received much less treatment. Nevertheless, while some of the cumulative variance expressions can be found in the literature (e.g., Ray, 1982), the core of the discussion here revolves around the covariance terms and the resulting Error Variance-Covariance matrix. We restrict our attention to basic ARIMA models, where the parameter order is fairly low, enabling relatively neat derivation of the expressions. Nonetheless, if the results hold for the basic ARIMA models, then they can be extended to more complex processes from the same family. Consequently, the following models are studied: ARIMA(0,0,0), ARIMA(0,1,0), ARIMA(0,0,1), ARIMA(0,1,1) and ARIMA(1,0,0). The inclusion of the first process, also known as i.i.d demand, will help explain the origin of the standard safety stock approximation and thus point to the inadequacy of its use for other processes differing from it.

For each process, it is presumed that the correct model and parameters are known a priori, implying that all forecasts are unbiased. We show that the correlation terms are present in these cases with no model mis-specification. If the results hold under this premise, then we postulate that they hold more generally. All forecasts are conditional on the information available at t. For ease of notation, y_t will represent demand at time t, and $Y_{t+L} = \sum_{i=1}^{L} y_{t+i}$, where

L is the lead time. The same logic applies for the conditional forecasts, $\widehat{Y}_{t+L|t} = \sum_{i=1}^{L} \widehat{y}_{t+i|t}$ and the forecast errors $e_{t+i|t} = y_t - \widehat{y}_{t+i|t}$. It should be noted that since the assumption of perfect demand knowledge is imposed here, then all forecasts are unbiased, and thus $\widehat{Y}_{t+L|t} = E[\widehat{Y}_{t+L|t}]$. The cumulative errors, $E_{t+L|t}$, which is the sum of errors up to lead time L, can be expressed as $E_{t+L|t} = Y_{t+L} - \widehat{Y}_{t+L|t} = \sum_{i=1}^{L} e_{t+i|t}$. The conditional variance of the cumulative errors may be written as

$$\operatorname{Var}\left(E_{t+L|t}\right) = \sum_{i=1}^{L} \operatorname{Var}\left(e_{t+i|t}\right) + 2 \sum_{i=1, i<j}^{L} \operatorname{Cov}\left(e_{t+i|t}, e_{t+j|t}\right). \tag{2.1}$$

If the errors are independent, then $\operatorname{Cov}\left(e_{t+i|t}, e_{t+j|t}\right) = 0$ for $i \neq j$, so the variance of the cumulative errors would just reduce to the sum of the individual variances and $\operatorname{Var}\left(E_{t+L|t}\right) = \sum_{i=1}^{L} \operatorname{Var}\left(e_{t+i|t}\right)$. However, despite the disturbances ε_t being independent, the forecasting errors e_t may display a correlation over time (Johnston and Harrison, 1986; Barrow and Kourentzes, 2016; Box et al., 2015), and this is demonstrated later in this section.

$\sum_{i=1,i<j}^{L} 2\operatorname{Cov}(e_{t+i|t}, e_{t+j|t})$, denotes the sum of all the covariances between the errors. This term differs for each process, depending on the DGP and on the forecast error structure. The conditional variance expressions for the forecast error in their variance-covariance matrix will be provided, which takes the form:

$$\underset{(L\times L)}{\Sigma} = \begin{bmatrix} \operatorname{Var}(e_{t+1|t}) & \operatorname{Cov}(e_{t+1|t}, e_{t+2|t}) & \operatorname{Cov}(e_{t+1|t}, e_{t+3|t}) & \dots & \operatorname{Cov}(e_{t+1|t}, e_{t+L|t}) \\ \operatorname{Cov}(e_{t+2|t}, e_{t+1|t}) & \operatorname{Var}(e_{t+2|t}) & \operatorname{Cov}(e_{t+2|t}, e_{t+3|t}) & \dots & \operatorname{Cov}(e_{t+2|t}, e_{t+L|t}) \\ \operatorname{Cov}(e_{t+3|t}, e_{t+1|t}) & \operatorname{Cov}(e_{t+2|t}, e_{t+3|t}) & \operatorname{Var}(e_{t+3|t}) & \dots & \operatorname{Cov}(e_{t+3|t}, e_{t+L|t}) \\ \dots & \dots & \dots & \dots & \dots \\ \operatorname{Cov}(e_{t+L|t}, e_{t+1|t}) & \operatorname{Cov}(e_{t+L|t+2|t}) & \operatorname{Cov}(e_{t+L|t+3|t}) & \dots & \operatorname{Var}(e_{t+L|t}) \end{bmatrix}$$

The Variance-Covariance Matrix Σ encompasses all the information required for the computation of the conditional variance of the errors for safety stocks. For example, the $L\sigma^2_{t+1|t}$ approach consists of using one element from the matrix, the top entry which corresponds to the one-step-ahead conditional variance, and multiplying it by the lead time. Given that the matrix is square, the sum of variances estimation method is its trace, $\operatorname{tr}(\Sigma)$. The true conditional variance of the cumulative errors at lead time L is the entry-wise sum of the Variance-Covariance matrix terms.

2.3.1 ARIMA(0,0,0)

The i.i.d demand model is a simple process and quite commonly assumed in the supply chain literature. This model is the ARIMA(0,0,0) with constant: $y_t = \mu + \varepsilon_t$. For a lead time of L, its $t+L$ actual value, forecast and forecast error expressions are $y_{t+L|t} = \mu + \varepsilon_{t+L}$, $\widehat{y}_{t+L|t} = \mu$ and $e_{t+L|t} = \varepsilon_{t+L}$, which gives $\text{Var}(e_{t+L|t}) = \sigma^2$. The term σ^2 refers to variance of the innovations. Usually, the variance of the forecast errors is not available before hand, and it has to be estimated conditionally on the data used. This estimated variance, $\widehat{\sigma}^2_{t+1|t}$, is different from the theoretical one, σ^2, and serves as a proxy for it. Nevertheless, in this section, since the focus lies on theoretical derivations, the estimated variance is not needed and the variance of the innovations is used instead. The above equations for the ARIMA(0,0,0) have no parameters apart from σ, nor do they include a recursive component on y_t, and as a result, no covariance terms appear. The forecast error term $e_{t+L|t}$ is always equated with the model innovation term ε_{t+L}, and so for this case the distribution of the forecast errors is equal to the distribution of the disturbances. For cumulative demand, $Y_{t+L|t} = L\mu + \sum_{i=1}^{L} e_{t+i|t}$ and $\widehat{Y}_{t+L} = L\mu$, and $E_{t+L|t} = \sum_{i=1}^{L} e_{t+i|t}$ and $\text{Var}(E_{t+L|t}) = L\sigma^2$. For any two integers $J < K$, the covariance between two errors is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \text{Cov}(\varepsilon_{t+J}, \varepsilon_{t+K}) = 0$. The Variance-Covariance Matrix is

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Given that $e_{t+i|t} = \varepsilon_{t+i}$ (assuming known parameters), the $L\sigma^2_{t+1|t}$ approximation is valid here. However, as we demonstrate next, the conditional variance of the cumulative errors adopts a different form for other DGPs, rendering the approximation inappropriate.

2.3.2 ARIMA(0,1,0)

The next model to be studied consists of the Random Walk model, ARIMA(0,1,0), or I(1); its demand is given by $y_t = y_{t-1} + \varepsilon_t$. A recursive term exists, inducing the appearance of the covariance terms, and a difference from the results pertaining to the i.i.d model. The

L-steps-ahead actual values are given by $y_{t+L} = y_t + \sum_{i=1}^{L} \varepsilon_{t+i}$, and the forecast $\hat{y}_{t+L|t} = y_t$. The corresponding error is $e_{t+L|t} = \sum_{i=1}^{L} \varepsilon_{t+i}$, and its L-steps-ahead variance of the errors is $\text{Var}(e_{t+L|t}) = L\sigma^2$. For a specified lead time L, the cumulative actuals, forecasts and errors are given as $Y_{t+L|t} = \sum_{i=1}^{L} y_{t+i} = Ly_t + \sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}$, and $\widehat{Y}_{t+L|t} = E_{t+L|t} = \sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}$. Their conditional variance is

$$\text{Var}(E_{t+L|t}) = \text{Var}\left(\sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}\right) = \left[\sum_{i=1}^{L} i^2\right]\sigma^2 = \sigma^2\left[\frac{L(L+1)(2L+1)}{6}\right]. \tag{2.2}$$

For any integers $J < K$, the covariance term is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = J\sigma^2$ (proof in Appendix A.1). Thus, the covariance of two terms is equal to the variance of the error with the smallest time index. The Variance-Covariance matrix for the error terms can be written as:

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \dots & \sigma^2 \\ \sigma^2 & 2\sigma^2 & 2\sigma^2 & \dots & 2\sigma^2 \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & \dots & 3\sigma^2 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & \dots & L\sigma^2 \end{bmatrix}$$

The Variance-Covariance matrix reveals that no entry is zero, and that they are all integer multiples of σ^2 for this process.

2.3.3 ARIMA(0,0,1)

The ARIMA(0,0,1) or MA(1) is written as $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$. It has short memory, retaining information from the previous shock only. To ensure the invertibility condition, we impose $|\theta| < 1$. Its L-steps-ahead actual value and forecasts are $y_{t+L} = \varepsilon_{t+L} + \theta\varepsilon_{t+L-1}$, and $\hat{y}_{t+L|t} = 0$. Its errors are $e_{t+L|t} = \varepsilon_{t+L} + \theta\varepsilon_{t+L-1}$, and its variance is $Var(e_{t+L|t}) = (1+\theta^2)\sigma^2$. For the L-steps-ahead cumulative demand, $Y_{t+L} = \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}$, and the forecast is $\widehat{Y}_{t+L|t} = \theta\varepsilon_t$. The error term is $E_{t+L|t} = \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}$, and its variance $\text{Var}(E_{t+L|t}) = \left[(L-1)(1+\theta)^2 + 1\right]\sigma^2$. If $J < K$, then: $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \theta\sigma^2$ if $K = J+1$ and 0 otherwise (see Appendix A.2). For this demand, only the consecutive errors are correlated, while the non-consecutive ones are uncorrelated. The Variance-Covariance Matrix is:

$$\Sigma = \begin{bmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 & 0 & ... & 0 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 & \theta\sigma^2 & ... & 0 \\ 0 & \theta\sigma^2 & (1+\theta^2)\sigma^2 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & (1+\theta^2)\sigma^2 \end{bmatrix}$$

In this case, the covariance terms can be negative and so can their sum, depending on the sign of θ.

2.3.4 ARIMA(0,1,1)

The ARIMA(0,1,1) is the well-known model underpinning Simple Exponential Smoothing, which is commonly applied in practical inventory management (e.g. Graves, 1999; Babai et al., 2013; Snyder et al., 2002, 2004). It is represented by $y_t = y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Similarly to the MA(1) process, the invertibility condition is imposed. The L-step-ahead actuals and forecasts are:

$$y_{t+L} = y_t + \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}, \tag{2.3}$$

and $\widehat{y}_{t+L|t} = y_t$. The L-steps-ahead forecast error is

$$e_{t+L|t} = \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}, \tag{2.4}$$

and its conditional variance of is $\text{Var}(e_{t+L|t}) = [1+(L-1)(1+\theta)^2]\sigma^2$. The L-step-ahead cumulative error is:

$$E_{t+L|t} = \varepsilon_{t+L} + \sum_{i=1}^{L-1}[1+(L-i+1)(1+\theta)]\varepsilon_{t+i} = \sum_{i=0}^{L-1}[1+i(1+\theta)]\varepsilon_{t+i+1} \tag{2.5}$$

with a conditional variance of

$$\begin{aligned} \text{Var}(E_{t+L|t}) &= \text{Var}\left(\sum_{i=0}^{L-1}[1+i(1+\theta)]\varepsilon_{t+i+1}\right) \\ &= \sigma^2 L\left[1+(1+\theta)(L-1)(\frac{(2L-1)(1+\theta)}{6}+1)\right] \end{aligned} \tag{2.6}$$

(proof in Graves (1999) or Babai et al. (2013)). For $J < K$, the covariance between two errors at different lead time is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \sigma^2(1+\theta)[1+(J-1)(1+\theta)]$ (proof in Appendix A.3). Only the J term appears in the equation, while K is absent. The Variance-Covariance matrix can be written as

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2(1+\theta) & \sigma^2(1+\theta) & \dots & \sigma^2(1+\theta) \\ \sigma^2(1+\theta) & [1+(1+\theta)^2]\sigma^2 & \sigma^2(1+\theta)[1+(1+\theta)] & \dots & \sigma^2(1+\theta)[1+(1+\theta)] \\ \sigma^2(1+\theta) & \sigma^2(1+\theta)[1+(1+\theta)] & [1+2(1+\theta)^2]\sigma^2 & \dots & \sigma^2(1+\theta)[1+2(1+\theta)] \\ \dots & \dots & \dots & \dots & \dots \\ \sigma^2(1+\theta) & \sigma^2(1+\theta)[1+(1+\theta)] & \sigma^2(1+\theta)[1+2(1+\theta)] & \dots & [1+(L-1)(1+\theta)^2]\sigma^2 \end{bmatrix}$$

2.3.5 ARIMA(1,0,0)

The ARIMA(1,0,0) or AR(1) model is a well-known demand process. Due to its simplicity, this model features in many inventory and supply chain research (e.g., Lee et al., 1997b; Chen et al., 2000a; Urban, 2005; Kahn, 1987). Its equation is $y_t = \phi y_{t-1} + \varepsilon_t$. The stationarity condition is imposed on the demand process to prevent such case, with $|\phi| < 1$. The $t+L$ expressions for the actuals and conditional forecasts are $y_{t+L} = \phi^h y_t + \sum_{i=1}^{L} \phi^{(L-i)} \varepsilon_{t+i}$, and $\widehat{y}_{t+L|t} = \phi^h y_t$. The L-step-ahead error is $e_{t+L|t} = \sum_{i=1}^{L} \phi^{L-i} \varepsilon_{t+i}$, and its variance $\text{Var}(e_{t+L|t}) = \sigma^2 \frac{(1-\phi^{2L})}{1-\phi^2}$. The L-steps-ahead cumulative actuals and forecasts are $Y_{t+L} = \sum_{i=1}^{L} y_{t+i} = \sum_{i=1}^{L} \sum_{j=1}^{L-i} \left[\phi^h y_t + \phi^{(L-j)} \varepsilon_{t+i}\right]$, and $\widehat{Y}_{t+L|t} = \phi^h y_t$. The error over lead time is $E_{t+L|t} = \sum_{i=1}^{L-1} \sum_{j=1}^{L-i} \left[(1+\phi^{(L-j)})\varepsilon_{t+i} + \varepsilon_{t+L}\right]$, with a variance of

$$\begin{aligned} \text{Var}(E_{t+L|t}) &= \text{Var}\left(\sum_{i=1}^{L-1}\sum_{j=1}^{L-i}(1+\phi^{(L-j)})\varepsilon_{t+i} + \varepsilon_{t+L}\right) \\ &= \frac{\sigma^2}{(1-\phi)^2}\sum_{i=1}^{L}(1-\phi^i)^2 \\ &= \frac{\sigma^2}{(1-\phi)^2}\left[L + \frac{1-\phi^{2L}}{1-\phi^2} - 2\left(\frac{1-\phi^L}{1-\phi}\right)\right] \end{aligned} \tag{2.7}$$

For $J < K$, the covariance between these the errors at these lead times is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \sigma^2 \phi^{(K-J)} \left[\frac{1-\phi^{2J}}{1-\phi^2}\right]$ (proof in Appendix A.4). Unlike the previous processes, $\text{Cov}(e_{t+J|t}, e_{t+K|t})$ depends on both J and K, as opposed to solely on J, and this arises due to the inclusion of the autoregressive parameter. The Variance-Covariance Matrix can be expressed as

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\phi & \sigma^2\phi^2 & ... & \sigma^2\phi^{L-1} \\ \sigma^2\phi & \sigma^2\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\phi\left[\frac{1-\phi^4}{1-\phi^2}\right] & ... & \sigma^2\phi^{L-2}\left[\frac{1-\phi^4}{1-\phi^2}\right] \\ \sigma^2\phi^2 & \sigma^2\phi\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\left[\frac{1-\phi^6}{1-\phi^2}\right] & ... & \sigma^2\phi^{L-3}\left[\frac{1-\phi^6}{1-\phi^2}\right] \\ ... & ... & ... & ... & ... \\ \sigma^2\phi^{L-1} & \sigma^2\phi^{L-2}\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\phi^{L-3}\left[\frac{1-\phi^6}{1-\phi^2}\right] & ... & \sigma^2\left[\frac{1-\phi^{2L}}{1-\phi^2}\right] \end{bmatrix}$$

2.3.6 Discussion of Variance-Covariance Expressions

The Variance-Covariance matrix derived for the DGPs encapsulate all the elements required for estimating the true lead time variance of the forecast errors. For processes other than the ARIMA(0,0,0), the $L\sigma^2_{t+1|t}$ is an inadequate mathematical expression for the lead time variance, and this can be seen from the L-steps-ahead variance expressions being greater than it (with the exception of the ARIMA(0,0,1) with a negative θ). Indeed, inspecting the Variance-Covariance Matrix for these DGPs reveals that the parameters also appear in the lead time variance, which are omitted nevertheless in the approximation. As mentioned earlier, this approximation only utilises the first entry of the matrix to scale the variance, and this proves to be insufficient for the estimation. The sum of variance approximation is inadequate as well, as it consists of summing up the diagonal entries of the matrix, assuming that the off-diagonal entries are zero. While it makes use of more elements from the matrix and thus an improvement over the previous approximation, it nevertheless is not enough. As can be seen from the different Variance-Covariance matrices, the off-diagonal entries are not necessarily zero, and these also render this approximation inappropriate, as the pairwise covariances are discarded from the estimation under the assumption that they are equal to zero. Both approximations will generally understate the true lead time variance for the forecast errors, as they ignore many components from the matrix that enter in its calculation. The sources of variance inflation are the pairwise covariance elements, which adopt a different form based on the model structure and are usually non-zero, as well as the different variance estimates at different lead times.

2.4 Approximations of Lead Time Error Variance

This chapter seeks to compare different approaches to approximating the variance of forecast errors over lead time. On a theoretical level the previous section showed that each demand model results in different covariance structures, and different values for the variance of errors, which depends on the demand process and its parameters. Additionally, if the parameters were estimated or the model was mis-specified, then we would expect the value of the covariance terms to be inflated as well. From a pragmatic viewpoint, the underlying DGP can assume many forms, and thus a parametric approach is not always possible. As a result, empirical approximations, independent of any assumptions of the demand process, are required. These methods approximate the lead time variance of errors from the observed forecasting errors only and do not rely on the forecasting model and its parameter. This section aims at introducing and discussing these approximations, and how they differ from each other.

2.4.1 Standard Approximation

As discussed in Section 2.2, the typical approximation is $L\widehat{\sigma}^2_{t+1|t}$, where $\widehat{\sigma}^2_{t+1|t}$ is often approximated by the in-sample MSE. This approximation has the limitations highlighted in section 2.2.2, but also is problematic due to the calculation of MSE, which may not represent the out-of-sample MSE, which is relevant to the inventory decisions being taken (Barrow and Kourentzes, 2016). The argument for this is that in-sample errors are naturally smaller than the forecast errors, as the forecasting method was fitted on that data and as such underestimates the observed conditional variance and ergo the future uncertainty (Makridakis and Winkler, 1989). Barrow and Kourentzes (2016) show that by switching to an out-of-sample estimate, using the MSE calculated on an appropriate validation set, the approximation becomes superior. Note that when a forecasting model is employed and estimated using Maximum Likelihood, the standard deviation of the innovation terms is an output of the estimation, which can differ from the least squares estimated value.

Alternatively, the one-step-ahead error can also be updated by a smoothing procedure (see for e.g. Syntetos and Boylan, 2006; Trapero et al., 2019a). In this work, the variance of the one-step-ahead forecast errors is calculated as:

$$\widehat{\sigma}^2_{t+1|t} = \frac{\sum_{t=1}^{n}\left(y_{t+1} - \widehat{y}_{t+1|t}\right)^2}{n} \tag{2.8}$$

with y_{t+1} and $\widehat{y}_{t+1|t}$ referring respectively to the realised demand and one-step-ahead conditional point forecasts.

2.4.2 Sum of Variances Approximation

Under this approximation, we calculate $\widehat{\sigma}^2_{t+1|t}$, $\widehat{\sigma}^2_{t+2|t}$,..., $\widehat{\sigma}^2_{t+L|t}$, with:

$$\widehat{\sigma}^2_{t+i|t} = \frac{\sum_{t=1}^{n}\left(y_{t+i} - \widehat{y}_{t+i|t}\right)^2}{(n-i)}, \tag{2.9}$$

and L is the desired lead time, and then sum them up to obtain the variance estimate $\sum_{i=1}^{L}\widehat{\sigma}^2_{t+i|t}$. This approximation falsely assumes that the errors at different steps-ahead are uncorrelated, but also suffers from a drawback due to the estimation of the multiple-steps-ahead MSE or $\widehat{\sigma}^2_{t+i|t}$. Given that multiple-steps-ahead errors are required, the forecasting model is fit on the in-sample data, and forecasts are generated using the rolling-origins procedure (Tashman, 2000). Subsequently, $\widehat{\sigma}^2_{t+i|t}$ is calculated on the multiple-step errors. Akin to the standard approximation, if only the training set is used, then the estimated variances may underestimate the future uncertainty, and out-of-sample data should be used as well. This can be achieved by using an appropriate validation set or recursively updating the estimates as new data points become available (Kourentzes, 2013). If the available sample is limited, an additional issue is that longer horizon error variance calculations may be based on very few errors, thus harming the quality of the estimation.

2.4.3 Cumulative Errors Approximation

While the ongoing discussion addresses the issues that are missed by the standard $L\widehat{\sigma}^2_{t+1|t}$ (and sum of variance approximation), it should be noted that it is not the norm for all researchers. For instance, Hyndman et al. (2008) provides the expression for the lead time variance of the exponential smoothing family of models, which can be used when fitting it to forecast demand in an inventory setting. Another method would consist of measuring the variance of errors over lead time, instead of calculating them at a horizon, and then aggregating them over lead time (Syntetos and Boylan, 2006; Trapero et al., 2019a). This empirical approximation is intuitive and computationally inexpensive, as it only entails calculating the cumulative forecast errors over lead time L, and then calculating its variance, or $\mathrm{Var}\left(\sum_{i=1}^{L} e_{t+i|t}\right) = \mathrm{Var}\left(\sum_{i=1}^{L}[y_{t+i} - \hat{y}_{t+i|t}]\right)$, where y_{t+i} denotes the actual value of demand at time $t+i$, and $\hat{y}_{t+i|t}$ its conditional forecast generated with information up until time t. This approximation is not new; for instance, Eppen and Martin (1988) employ this heuristic when calculating the variance of errors, but do not provide the rationale behind its use. Lee (2014) uses this method for estimating the variance of the errors, justifying the choice as it captures the uncertainty in lead time demand, but not explaining how. More recently, Trapero et al. (2019a) also argue that using this approach offers the benefits of not needing to specify a model or its parameters. This method of tracking forecast errors seems a natural path to follow as it covers all the errors made during L; furthermore, since it contains all the information of the distribution of cumulative errors, it should mitigate the impact of the above mentioned correlations. Nevertheless, the motivation behind it has not been clearly defined in the literature. The variance-covariance matrix shown in section 2.3 exhibits all the components of the total variance of errors over lead time, which are featured in the cumulative error approximation. Given that the cumulative errors are comprised of the forecast errors aggregated over lead time L, we expect this approach to smooth out the errors, similarly to any other aggregation process. Finally, with the aggregation over L, we also expect the Central Limit Theorem will render the errors distributions closer to Normal, especially with higher values of the lead time (Silver et al., 2016).

The three approximations for estimating the error variance, which were elaborated in this section, will be compared via simulation in the next section, to assess their performance in an inventory context.

2.5 Inventory Simulation

So far, the theoretical discussion implies that even with perfect knowledge of the demand process, the errors at different lead times are correlated, and these correlations accrue in the computation of the conditional variance of the aggregate errors over lead time. Any effect on the safety stock has to account for the inventory policy. To evaluate the losses due to omitting the aforementioned covariances and the performance of the various approximations, we employ an inventory simulation.

2.5.1 Experimental Setup

The objective of the simulation is to compare the inventory performance of the three alternatives, outlined in section 2.4, namely: (i) the regular that uses the $L\widehat{\sigma}^2_{t+1|t}$; (ii) the sum of error variances; and (iii) the cumulative error variance.

Following the theoretical discussion, we consider six DGPs, as shown in table 2.2. Three are stationary patterns (AR, MA and ARMA) and three are non-stationary so as to investigate both cases. While four of the above processes featured in the analytical discussion in section 2.3, the ARMA and ARIMA model allow the inclusion of processes with both autoregressive and moving average features of low order. The model innovations ε_t are assumed to be normally distributed, with their variance, σ, taking the values: $\{1,5,10,25\}$, to represent different levels of demand volatility, and the parameters for each model are drawn from a Uniform distribution, with $0 < \phi, \theta < 1$ to impose the condition of stationarity and invertibility for each process (Box et al., 2015). For all processes we add a constant level to guarantee positive values of demand. The full set of control parameters for the experimental design are listed in the table below.

Variable	Values	Options
Variance Approximations	Regular, Cumulative, Sum	3
Forecasting Models (Uncertainty Type)	Sample Size (SSU), Parameter (PU), Model (MU)	3
Demand Process (Stationary)	AR(1), MA(1), ARMA(1,1,)	3
Demand Process (Nonstationary)	I(1), IMA(1,1), ARIMA(1,1,1)	3
Demand Noise Level	1, 5, 10, 25	4
Retailer Horizon	3, 5	2
Retailer Service Level	85%, 90%, 95%	3

Table 2.1: Experimental Design Control Parameters

Table 2.2: Simulation DGPs

	DGP	Formula
Stationary	AR(1)	$Y_t = \phi Y_{t-1} + \varepsilon_t$
	MA(1)	$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$
	ARMA(1,1)	$Y_t = \phi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$
Non-stationary	I(1)	$Y_t = Y_{t-1} + \varepsilon_t$
	IMA(1,1)	$Y_t = Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$
	ARIMA(1,1,1)	$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$

For each series we generate 400 observations and split it into three subsets: the first 100 points are used as training set, where the forecasting model is fitted and the parameters are estimated; the following 200 observations are used as "burn-in", i.e. the inventory policy is allowed to run over these to eliminate the impact of any initial inventory settings (Kourentzes, 2013); and the last 100 observations constitute the test set, where the experiment outputs are recorded and results are computed. We generate 500 replications for each DGP.

We consider the three aforementioned uncertainties: (i) Sample Size Uncertainty (SSU), where the demand model and parameters are known a priori and only σ is unknown; (ii) Parameter Uncertainty (PU) where the DGP formula is known but the parameters need to be estimated via Maximum Likelihood Estimation (MLE), and (iii) Model Uncertainty (MU), where nothing is known about the DGP and an incorrect forecasting model is deliberately fit. Each of the three uncertainty scenarios magnifies potential problems in the estimation of the variance and therefore we can assess the impact on the inventory outcome. These are reflected in three forecasting models being fitted to each DGP, each corresponding to an uncertainty scenario.

An Order-Up-To (R,S) periodic inventory policy is considered, with a review period $R = 1$ and the measured horizons are $L+R = \{3,6\}$. The horizons are assumed to be deterministic in order to isolate any of their variability on the total variability of safety stocks. We do not consider the one-step ahead case, as none of the discussed covariances will appear. When an out-of-stock occurs any sales are lost, we assume that they are satisfied by a competitor. Given the difficulty in calculating the costs associated with stock-outs and lost sales, this chapter will focus on setting safety stocks based on a customer service level. We track the performance at $85\%, 90\%$ and 95% target service levels, to better understand the impact of the latter on the inventory implications of each approximation. This is done by means of trade-off curves between lost sales and inventory-on-hand (Gardner, 1990), as well as realised α-cycle service level.

For each forecasting model representing an uncertainty scenario, the multiple-steps ahead point forecasts are produced following a rolling-origins scheme, and then aggregated over $L+R$ to obtain the cumulative forecasts. The safety stocks are calculated according to the estimation methods explained in section 2.4 for each type of approximation as:

$$SS_t = k_\alpha \widehat{\sigma}_t^M \tag{2.10}$$

where SS_t denotes the safety stocks at time t, k the the inverse of the normal distribution associated with the desired cycle service level α, and $\widehat{\sigma}_t^M$ the standard deviation of the forecast errors over $R+L$ for approximation M. The appropriate quantiles are constructed based on an empirical estimation of the conditional variance of the errors for all forecasting models, since the goal of this simulation is to evaluate the performance of the approximations. To determine the conditional variance of the error term, equation 2.9 is employed, where we use the errors in the out-of-sample subset ("burn-in" and "test" sets in our case) as it becomes available, instead of the in-sample errors, for the reasons discussed in section 2.4. As the inventory simulation progresses, any new forecast errors are included in the calculation. The simulation was conducted using the R statistical language (R Core Team, 2019).

2.5.2 Simulation Results

2.5.2.1 Comparison of the three approximations

Analysing the results reveals that different σ values for the model innovations do not affect the findings greatly, and therefore we report only for $\sigma = 10$. We group the presentation of the results into the stationary and the non-stationary cases (see table 2.2), as they exhibit similar behaviour. Table 2.3 reports the deviations of service level between realised and target, for the different horizons $L+R$, across simulation repetitions and DGPs. These are visualised in Figure 2.1. We calculate the difference from the realised service levels so that the direction of the deviations stays the same as the observed service levels, i.e. positive deviations imply larger than the target service levels. It should be noted that since service level deviations are a relative metric rather than an absolute one, we are interested in how close its magnitude is with respect to 0.

Consistently across target service levels and types of uncertainty the cumulative estimation returns the highest service levels, followed by the sums and finally the regular. This is due to the underlying estimate of the error variance, which is ranked in the same order. Although this follows from the theoretical discussion, in section 2.3, for the stationary processes the higher service levels do not necessarily result in lower service levels deviations. At $L+R=3$, the cumulative estimation has overall the lowest deviation from the target. Notably, as the uncertainty increases (with MU having the maximum), the deviations of the cumulative do not suffer, in contrast to the other approaches. For the regular and sum approaches, while for SSU and PU any differences in the results are minimal, the performance worsens rapidly for the MU case. Note that given the adequate estimation sample size, it is expected that there will be only small errors in the estimation of the model parameters, therefore making SSU and PU exhibit similar results. As the lead time increases ($L+R=6$) the error variance increases as well. This is to be expected, as forecasting longer horizons is more challenging. However, when this is translated to service levels, it causes a global shift of the realised service levels upwards. This causes the cumulative approach to overshoot its target, resulting in additional stock holding than what was targeted. Noticeably, the regular approach typically under-performs in terms of meeting the cycle service level, and either sum of variances or the cumulative approaches are always preferable for achieving this end. A final useful observation is that in all cases, for the stationary time series the range of service level deviations is $[-8.77\%, 5.35\%]$, with the overshoots by the cumulative approach being quite small.

Shifting our attention to the service level deviations for the non-stationary process, provided in Table 2.4 and Figure 2.2, we can observe that the results are ordered in the same fashion. However, now in all cases the deviations are negative. The cumulative approach, for both $L+R=3$ and $L+R=6$, results in the minimal deviations, ranging from $[-8.24\%, -0.79\%]$ that although consistently negative, are of similar scale to the deviations observed for the stationary series. This performance is followed by the sum of variances, which again exhibits

Uncertainty	Safety Stock	$\alpha = 85\%$		$\alpha = 90\%$		$\alpha = 95\%$	
		$L = 3$	$L = 6$	$L = 3$	$L = 6$	$L = 3$	$L = 6$
SSU	Regular	-2.76%	**-0.24%**	-5.18%	-3.31%	-6.89%	-5.74%
	Sum	**-1.15%**	1.36%	-3.31%	**-1.45%**	-4.87%	-3.64%
	Cumulative	2.48%	5.32%	**0.55%**	2.76%	**-0.92%**	**0.65%**
PU	Regular	-2.84%	**-0.49%**	-5.18%	-3.57%	-6.95%	-6.01%
	Sum	**-1.20%**	1.11%	-3.36%	**-1.73%**	-4.81%	-3.79%
	Cumulative	2.56%	5.35%	**0.68%**	2.93%	**-0.79%**	**0.72%**
MU	Regular	-4.31%	-1.91%	-6.94%	-5.30%	-8.77%	-8.14%
	Sum	-3.78%	**-1.16%**	-6.35%	-4.48%	-8.09%	-7.17%
	Cumulative	**2.19%**	5.25%	**0.29%**	**2.72%**	**-1.09%**	**0.56%**

Table 2.3: Achieved α service level deviations for stationary demand processes

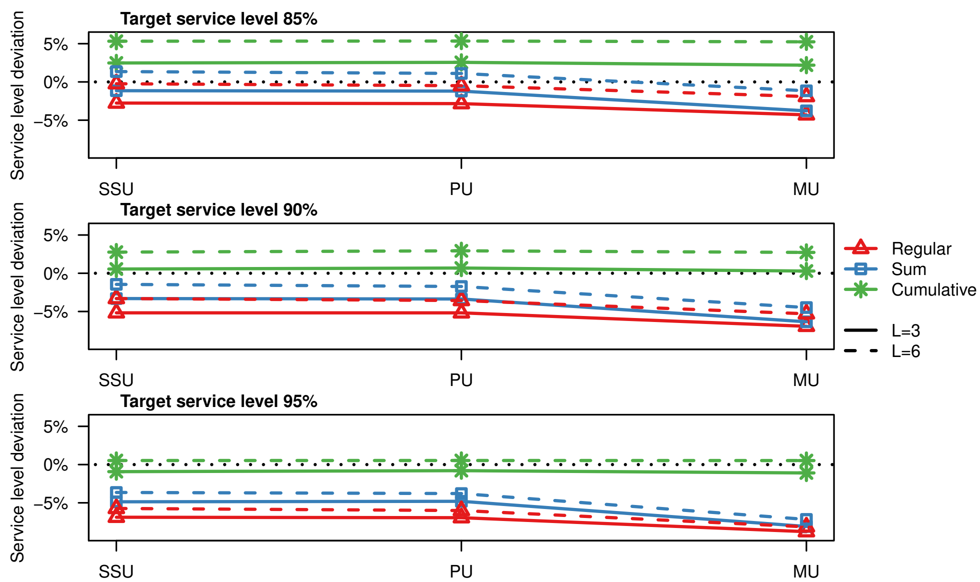

Figure 2.1: α-Service levels deviations for stationary demand processes

large drops under full model uncertainty, that is the typical case in practice. The regular approach always performs worse, following our understanding from the theoretical discussion. Note that the deviations for the non-stationary processes, for the sum and regular approaches range between $[-24.09\%, -5.93\%]$, which is interesting to contrast with the observed range for the cumulative and the results for the stationary processes. The latter comparison demonstrates that even if in the stationary case the cumulative approach demonstrated a tendency to overshoot its target, that deviation was comparatively very small.

The service levels deviations tell only part of the full story. We refer to Figures 2.3 and 2.4 for the trade-off curves. In both stationary and non-stationary cases we observe that

Uncertainty	Safety Stock	$\alpha = 85\%$		$\alpha = 90\%$		$\alpha = 95\%$	
		$L = 3$	$L = 6$	$L = 3$	$L = 6$	$L = 3$	$L = 6$
SSU	Regular	-11.86%	-9.01%	-15.67%	-13.54%	-19.03%	-17.81%
	Sum of Variances	-8.38%	-5.96%	-11.55%	-9.80%	-13.88%	-13.13%
	Cumulative	**-4.32%**	**-0.84%**	**-6.71%**	**-3.80%**	**-8.24%**	**-6.09%**
PU	Regular	-11.82%	-9.01%	-15.62%	-13.51%	-18.97%	-17.77%
	Sum of Variances	-8.37%	-5.93%	-11.52%	-9.76%	-13.86%	-13.11%
	Cumulative	**-4.29%**	**-0.79%**	**-6.67%**	**-3.74%**	**-8.22%**	**-6.04%**
MU	Regular	-16.06%	-15.56%	-19.95%	-19.99%	-23.27%	-24.09%
	Sum of Variances	-14.92%	-14.14%	-18.66%	-18.19%	-21.77%	-21.89%
	Cumulative	**-6.77%**	**-4.76%**	**-8.03%**	**-6.22%**	**-7.93%**	**-6.35%**

Table 2.4: Achieved α service level deviations for non-stationary demand processes

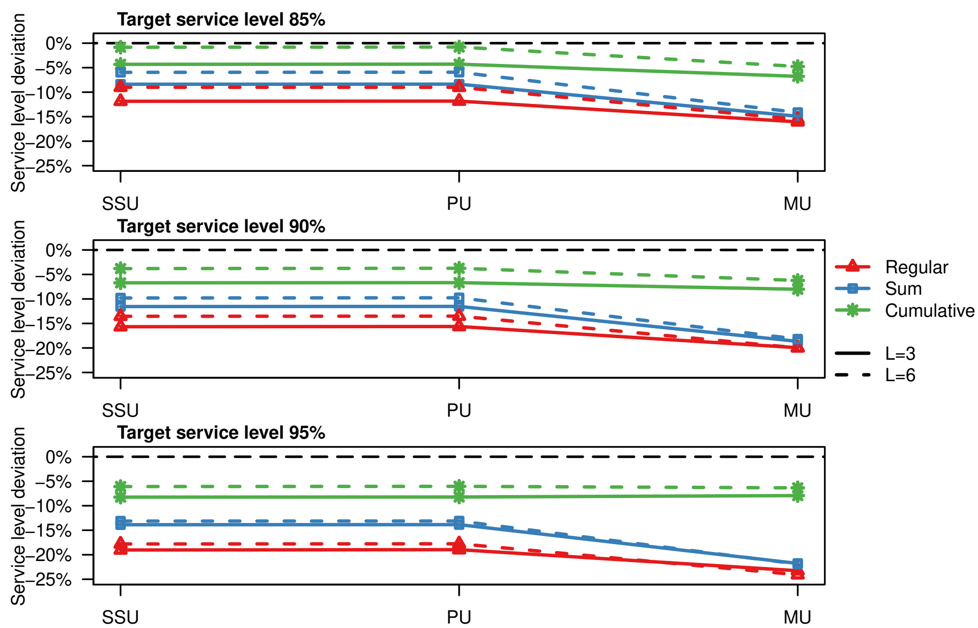

Figure 2.2: α-Service levels deviations for non-stationary demand processes

in terms of lost sales the regular approach returns the highest value, while the cumulative has the lowest. Furthermore, the excess inventory follows the expected outcomes, with approaches that exhibit the least lost sales naturally retaining more stock. It is interesting to observe that for a given type of uncertainty and lead time, no approach results in a trade-off curve that clearly dominates the others, only shifting the curves to a different balance point between lost sales and excess stock. Pairing this with the service levels deviations values discuss above, the better performance of the cumulative approach becomes evident. When the horizon increases from 3 to 6 periods, all trade-off curves are shifted to the right, i.e. a higher stock position. Under model uncertainty the good performance of the cumulative approach

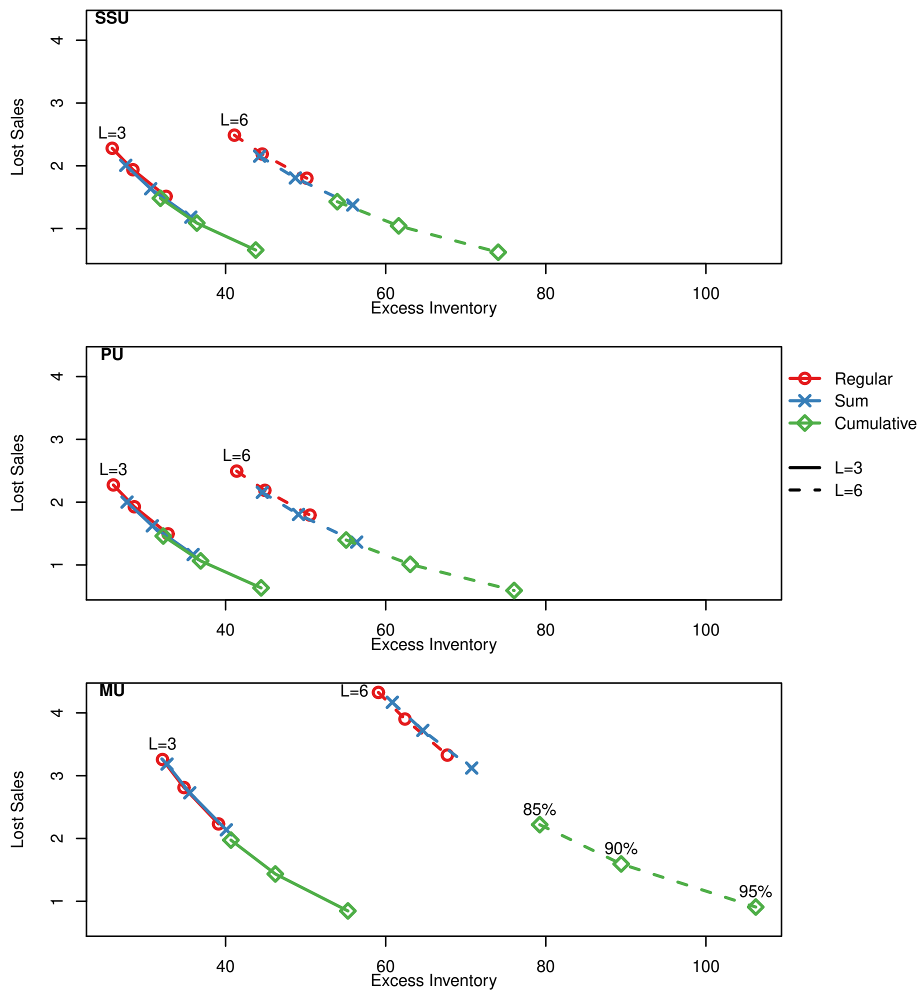

Figure 2.3: Trade-off curves for stationary demand processes (by lead time and uncertainty type). The values displayed on the vertical axis denote the total lost sales over the test set, while those on the horizontal axis represent the total excess inventory for the same period. The correspondence of the curves to the target service levels (85%, 90% and 95%) is provided for one instance for the curve on the right, corresponding to cumulative approach with lead time 6.

is further highlighted, as it is the only one that does not incur a substantial increase in lost sales, irrespective of the target service level.

Considering the non-stationary time series (Figure 2.4), we get a similar insight in the performance of the competing approaches. However, the scale of both lost sales and excess stock is increased, reflecting the difficulty in forecasting these processes. As the modeling uncertainty increases, both the regular and the sum of variances approaches deteriorate rapidly, echoing the results provided for the service levels. Comparing the curves for $L+R=3$ and

$L = R = 6$, we observe that the latter lies further on the right, which implies that more excess inventory is incurred (which is to be expected as the lead time increases). In addition, we also notice an upward shift from $L + R = 3$ to $L = R = 6$ in lost sales, which leads us to conclude that both approximations display a poor performance. On the other hand, the cumulative approach does not exhibit a substantial increase in lost sales, demonstrating the need to account for the covariances discussed in the theoretical part.

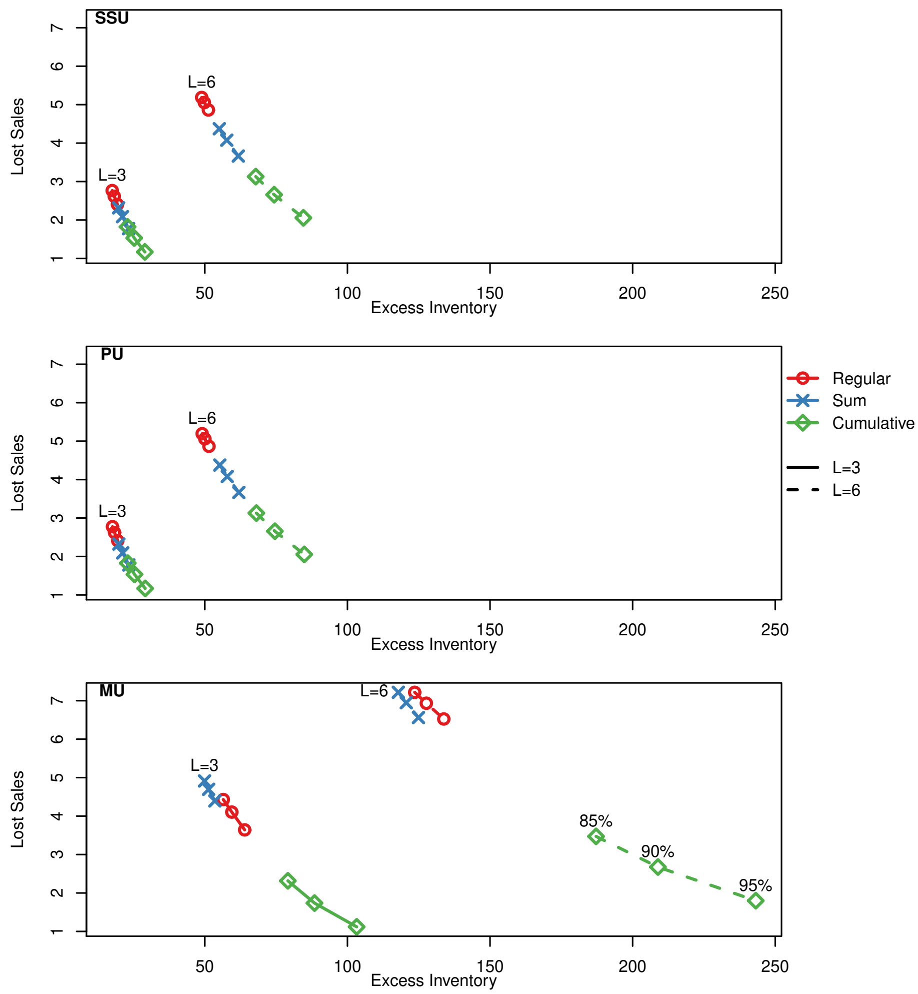

Figure 2.4: Trade-off curves for non-stationary demand processes (by lead time and uncertainty type). The correspondence of the curves to the target service levels (85%, 90% and 95%) is provided for one instance.

Overall, we find that across all cases, i.e. process type, uncertainty type and lead time, on average the cumulative approach performs best, followed by the sum of variance and the regular ranking the worst. The sum of variances captures better the long-term forecast error variances, but ignores the additional covariance terms that are reflected in the calculation procedure of the cumulative approach. The latter exhibits the minimal deviation from the target service level, which are consistently small, across all cases. This is achieved without shifting to a dominated trade-off curve, that is it does not need to accumulate unreasonably excess stock and simply results in an appropriate balance between lost sales and excess stock. The other approaches do not achieve this due to the omitted covariances. Its superior performance becomes even more attractive given its implementation simplicity.

2.6 Real Data Study

We validate the simulation insights by testing the competing variance approximation methods on a real dataset. We use sales data of dairy and cheese products from an American retailer, containing 111 daily series ranging from 521 to 2860 observations.[1] Compared to the simulated DGPs, many of these time series exhibit promotional effects and seasonality, which would be particularly prone to Model Uncertainty. Similar to before, we use an (R,S) inventory policy, with the same service levels and lead times, and a threefold partitioning of the data. The "burn-in" period and test sets consist of 75 observations each, and the remainder is used to estimate the forecasting model parameters. This can be any exponential smoothing model, as selected by minimising the AICc criterion (for details see Hyndman and Khandakar, 2008). Naturally, even with model selection, this setting falls under the model uncertainty case, as the true underlying data process is unknown. We acknowledge that in practice the retailer might employ a different inventory policy, given the software's limitations, the nature of the product and the sampling frequency of the data collected; however the use of this dataset is dictated by our experimental requirements rather than mirroring the retailing environment.

The results are presented in Figures 2.5 and 2.6 that present the service level deviations and trade-off curves respectively. The service level deviations range lies between $[-7.55\%, 7.1\%]$. A tendency to over-cover can be observed, especially for the lowest service level of 85% or high lead times, which can be attributed to the presence of promotions that

[1] The dataset used is that of *Dominick's Finer Food*, made publicly available by the University of Chicago at : http://research.chicagogsb.edu/marketing/databases/dominicks/

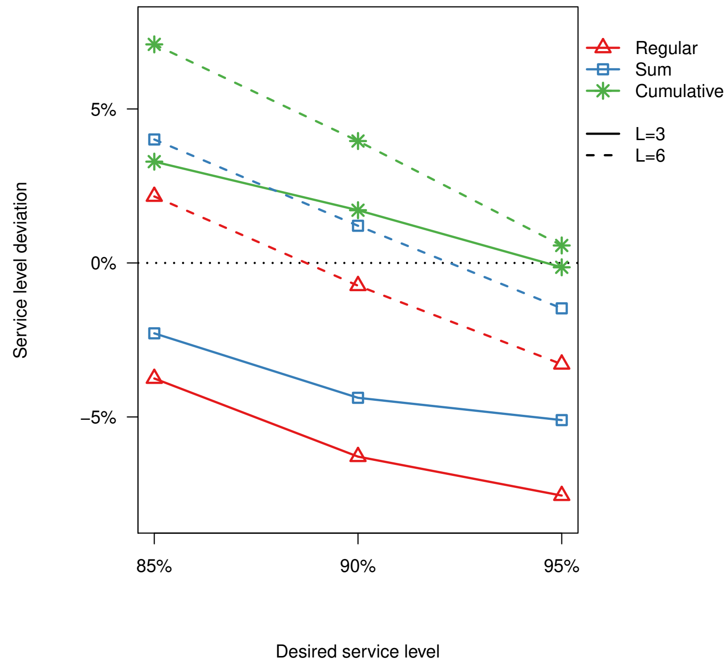

Figure 2.5: α-Service levels deviations for case study data

result in higher forecast error variance estimates, thus further exacerbating the model uncertainty component. Overall, the performance of the cumulative errors is better than the other methods. This is particularly evident for the 95% service level, where its deviations lie very closely to the optimal deviation line, as opposed to the other two methods that return higher deviations.

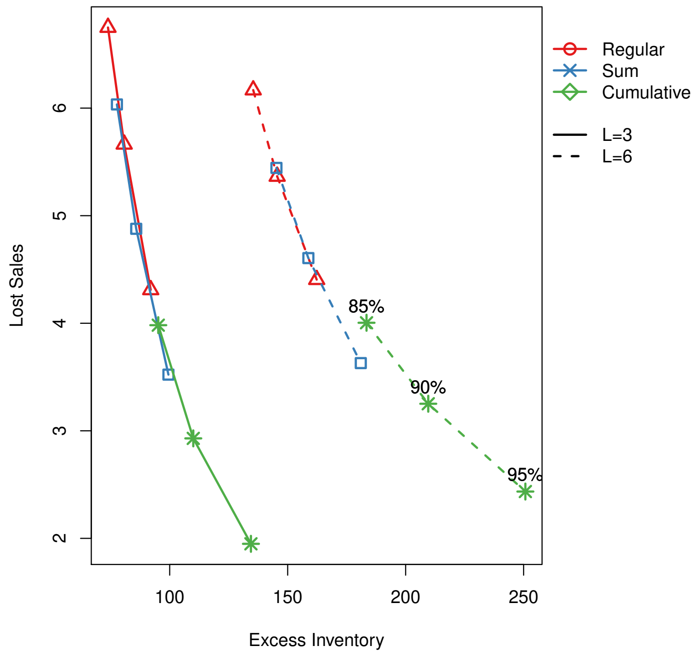

Figure 2.6: Trade-off curves for case study data

The trade-off curves in Figure 2.6 reveal a similar story to those of the simulated data, as the cumulative errors return lower levels of lost sales at the cost of more stock on hand. Again, similarly to the simulation findings, no approximation dominates both aspects of minimising lost sales and excess inventory. But the important result from the graph is the progression of the deviations as the service level increases. Indeed, at higher service levels (which are the norm in practice), the cumulative approach tends to get close to the dashed horizontal line corresponding to zero deviations, which is the ideal scenario, while the other two methods, as they have negative deviations, further diverge from the zero line. The regular approximation, on the other hand, displays the highest level of stock-outs. No set of curves is clearly dominated by others, and when this is paired with the service level deviations, we conclude that the cumulative variance approximation achieves a preferable trade-off between lost sales and stock on hand.

2.7 Conclusion

Safety stocks are a key element for many business operations, as they allow decision makers to alleviate the effect of demand uncertainty on order levels. A critical component of safety stocks lies in the estimation of the variance of forecast errors, which is used to quantify the uncertainty surrounding it. When computing this variance, the correlation between forecast errors at different lead times is often ignored, which results in lower estimates, which in turn generate sub-par ordering decisions. The existence of these correlations stems from the underlying demand model structure and was demonstrated for simple demand processes even under full knowledge of the generating scheme. This chapter has examined this issue in a twofold fashion by: (i) pointing out the shortcomings of the standard heuristic employed by many, and (ii) arguing for the use of cumulative errors, an easy and straightforward empirical approximation, which takes advantage of the distribution of the forecast errors over lead time.

The theoretical expressions demonstrated that for the demand processes investigated, the $L\widehat{\sigma}^2_{t+1|t}$ method is not suited to approximate the variance of demand, as it will understate it. This was investigated by deriving the exact lead time forecast error variances of simple ARIMA models, and comparing these expressions with the standard approximation. The Monte Carlo simulation indicated the superiority of the proposed heuristic over the traditional one for an Order-Up-To inventory policy with deterministic lead times. These results held for three types of demand uncertainty. Furthermore, cumulative errors seemed to handle model uncertainty, the typical case in practice, better than its counterparts. These findings were supported by similar ones for a case study using a real dataset from a retailer.

This chapter has shown that using cumulative errors, while being an empirical approximation, is more appropriate for finding the variance of forecast errors over lead time. While not being a new method, this work provides the motivation to prefer it over more common approaches and provides an inventory based evaluation of its performance. The findings are of particular interest to practitioners, who can fairly easily adopt the approximation, which is particularly effective when the underlying DGP is unknown, reflecting reality. Its ease of implementation and the absence of the need to specify any parameters or variance expressions make it attractive, as it can be easily implemented in existing inventory software.

A Derivation of the Covariance Expressions

Recall that $\text{Cov}(aU + bV, cY + dZ) = ac\text{Cov}(U,Y) + bc\text{Cov}(U,Y) + ad\text{Cov}(V,Y) + bd\text{Cov}(V,Z)$, with a, b, c, d constants, and U, V, Y, Z variables. Given

$$\begin{cases} \text{Cov}(\varepsilon_{t+i}, \varepsilon_{t+i}) = \text{Var}(\varepsilon_{t+i}) = \sigma^2 \\ \text{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}) = 0, \quad i \neq j, \end{cases}$$

and since the error $e_{t+L|t}$ is a linear function of the model innovations up to lead time L, the covariances can be computed by expanding the former in order for the innovations terms to appear in the estimation.

A.1 ARIMA(0,1,0)

For any integers J and K, such that $J < K$, we have

$$e_{t+K|t} = \sum_{i=1}^{K} \varepsilon_{t+i} = \sum_{i=1}^{J} \varepsilon_{t+i} + \sum_{i=J+1}^{K} \varepsilon_{t+i} = e_{t+J|t} + \sum_{i=J+1}^{K} \varepsilon_{t+i}. \tag{11}$$

Because of the model structure, $e_{t+K|t}$ can be written as a sum of $e_{t+J|t}$ and other succeeding innovations. For $J < K$, we have

$$\begin{aligned} \text{Cov}\left(e_{t+J|t}, e_{t+K|t}\right) &= \text{Cov}\left(e_{t+J|t}, e_{t+J|t} + \sum_{i=J+1}^{K} \varepsilon_{t+i}\right) \\ &= \text{Cov}\left(e_{t+J|t}, e_{t+J|t}\right) + \sum_{i=J+1}^{L} \text{Cov}\left(e_{t+J|t}, \varepsilon_{t+i}\right) \\ &= Var\left(e_{t+J|t}\right) + 0 \\ &= \text{Var}\left(\sum_{i=1}^{J} \varepsilon_{t+i}\right) \\ &= J\sigma^2 \end{aligned} \tag{12}$$

A.2 ARIMA(0,0,1)

For two integers $J < K$, we have:

$$\begin{aligned}
\mathrm{Cov}(e_{t+J|t}, e_{t+K|t}) &= \mathrm{Cov}(\varepsilon_{t+J} + \theta\varepsilon_{t+J-1}, \varepsilon_{t+K} + \theta\varepsilon_{t+K-1}) \\
&= \mathrm{Cov}(\varepsilon_{t+J}, \varepsilon_{t+K}) + \theta\mathrm{Cov}(\varepsilon_{t+J-1}, \varepsilon_{t+K}) \\
&+ \theta\mathrm{Cov}(\varepsilon_{t+K-1}, \varepsilon_{t+J}) + \theta^2\mathrm{Cov}(\varepsilon_{t+K-1}, \varepsilon_{t+J-1}).
\end{aligned} \tag{13}$$

If $K \neq J+1$ then the covariance terms all reduce to 0. If however $K = J+1$, then

$$\begin{aligned}
\mathrm{Cov}(e_{t+J+1|t}, e_{t+J|t}) &= \mathrm{Cov}(\varepsilon_{t+J+1}, \varepsilon_{t+J}) + \theta\mathrm{Cov}(\varepsilon_{t+J+1}, \varepsilon_{t+J-1}) \\
&+ \theta\mathrm{Cov}(\varepsilon_{t+J+1-1}, \varepsilon_{t+J}) + \theta^2\mathrm{Cov}(\varepsilon_{t+J-1}, \varepsilon_{t+J-1}) \\
&= \theta\mathrm{Var}(\varepsilon_{t+J+1-1}) \\
&= \theta\mathrm{Var}(\varepsilon_{t+J}) \\
&= \theta\sigma^2
\end{aligned} \tag{14}$$

A.3 ARIMA(0,1,1)

For two integers $J < K$, $e_{t+K|t} = \theta\varepsilon_t + \sum_{i=1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}$ $e_{t+J|t} = \theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+J}$, and $e_{t+K|t} = \theta\varepsilon_t + \sum_{i=1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}$. This implies that

$$\begin{aligned}
e_{t+K|t} &= \theta\varepsilon_t + \sum_{i=1}^{J}(1+\theta)\varepsilon_{t+i} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K} \\
&= \theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + (1+\theta)\varepsilon_{t+J} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K} \\
&= e_{t+J|t} + \theta\varepsilon_{t+J} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}
\end{aligned} \tag{15}$$

The covariance between the errors at these lead times is

$$\mathrm{Cov}(e_{t+J|t}, e_{t+K|t}) = \mathrm{Cov}\left(e_{t+J|t}, e_{t+J|t} + \theta\varepsilon_{t+J} + \sum_{i=J+1}^{K-1}\varepsilon_{t+i} + \varepsilon_{t+K}\right). \tag{16}$$

Since $e_{t+J|t}$ is a combination of ε_{t+i} terms up until J, and $J < K$, then $\mathrm{Cov}(e_{t+J|t}, \sum_{i=J+1}^{K-1}\varepsilon_{t+i}) =$

0. The same applies for $\text{Cov}(e_{t+J|t}, \epsilon_{t+K}) = 0$. Therefore, the covariance becomes

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+K|t}) &= \text{Cov}(e_{t+J|t}, e_{t+J|t} + \theta\varepsilon_{t+J}) \\ &= \text{Cov}(e_{t+J|t}, e_{t+J|t}) + \theta\text{Cov}(e_{t+J|t}, \varepsilon_{t+J}).\end{aligned} \tag{17}$$

The second term in the equation amounts to

$$\begin{aligned}\text{Cov}(e_{t+J|t}, \varepsilon_{t+J}) &= \text{Cov}\left(\theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+J}, \varepsilon_{t+J}\right) \\ &= \text{Cov}(\varepsilon_{t+J}, \varepsilon_{t+J}) \\ &= \sigma^2.\end{aligned} \tag{18}$$

And

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+J|t}) &= \text{Var}(e_{t+J|t}) \\ &= \sigma^2[(J-1)(1+\theta)^2 + 1],\end{aligned} \tag{19}$$

Implying

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+K|t}) &= \text{Var}(e_{t+J|t}) + \theta\sigma^2 \\ &= \sigma^2[1 + (J-1)(1+\theta)^2 + \theta] \\ &= \sigma^2\left[(1+\theta) + (J-1)(1+\theta)^2\right] \\ &= \sigma^2(1+\theta)[1 + (J-1)(1+\theta)]\end{aligned} \tag{20}$$

This is the covariance between the errors at different lead times.

A.4 ARIMA(1,0,0)

For $J < K$, the error at lead time J is $e_{t+J|t} = \sum_{i=1}^{J} \phi^{(J-i)}\varepsilon_{t+i}$, while that at lead time K is

$$\begin{aligned}e_{t+K|t} &= \sum_{i=1}^{K} \phi^{(K-i)}\varepsilon_{t+i} = \sum_{i=1}^{K} \phi^{(K-J+J-i)}\varepsilon_{t+i} \\ &= \phi^{(K-J)}e_{t+J|t} + \sum_{i=J+1}^{K} \phi^{(K-i)}\varepsilon_{t+i}.\end{aligned} \tag{21}$$

The covariance between the errors at these lead times is

$$
\begin{aligned}
\operatorname{Cov}(e_{t+K|t}, e_{t+J|t}) &= \operatorname{Cov}\left(e_{t+J|t}, \phi^{(K-J)} e_{t+J|t} + \sum_{i=J+1}^{K} \phi^{(J-i)} \varepsilon_{t+i}\right) \\
&= \operatorname{Cov}\left(\phi^{(K-J)} e_{t+J|t}, e_{t+J|t}\right) \\
&+ \operatorname{Cov}\left(\phi^{(K-J)} e_{t+J|t}, \sum_{i=J+1}^{K} \phi^{(J-i)} \varepsilon_{t+i}\right) \\
&= \phi^{(K-J)} \operatorname{Var}(e_{t+J|t}) \\
&= \sigma^2 \phi^{(K-J)} \left[\frac{1-\phi^{2J}}{1-\phi^2} \right]
\end{aligned}
\tag{22}
$$

Chapter 3

Forecast Uncertainty as a Determinant for the Bullwhip Effect

The Bullwhip Effect, defined as the magnification of demand variability throughout the supply chain, poses a challenge to many firms. One of its contributing factors is inaccurate forecasts used to support demand planning. Forecast errors translate into higher inventory costs at a local level, but also impact other members in the supply chain as their decisions are based on the mis-estimated incoming orders. The conventional measure for the Bullwhip Effect does not reflect how forecast uncertainty evolves in the supply chain. This chapter proposes a new metric to serve this purpose, which overcomes many of the limitations of the Bullwhip Ratio: the Ratio of Forecast Uncertainty, which benchmarks the upstream Mean Squared Errors of forecasts aggregated over lead time to the downstream's. To study the properties and relevance of this measure, an inventory simulation is deployed. Our results indicate that this metric bears a connection to inventory costs at upstream level, and holds more explanatory power than the standard Bullwhip Ratio. Managers can make use of this ratio to better understand the upstream impact of the the forecasting process in the entire supply chain and the cost-saving gains achieved by improving it.

3.1 Introduction

A problem faced by many supply chains is the so-called "Bullwhip Effect". This is caused by the distortion of information as it moves upstream in the supply chain, and manifests itself with an increase in the variability of orders. The Bullwhip Effect poses an important challenge to many supply chains, as it causes firms to bear unnecessary costs and experience higher levels of stockouts and customer dissatisfaction. Despite similar findings being detected in earlier studies (Forrester, 1961; Sterman, 1989), it was the seminal paper by Lee et al. (1997a) that drew the attention of academics to this matter (Wang and Disney, 2016). Its existence has been verified empirically at different industry levels (Trapero et al., 2012; Zotteri, 2013; Shan et al., 2014; Isaksson and Seifert, 2016; Pastore et al., 2019a). Ever since, many studies have been dedicated to this topic, addressing its different causes and facets, examining the impact of various factors on it, and suggesting remedies to counter and reduce the phenomenon.

One of the established causes for the Bullwhip Effect is demand signal processing (Lee et al., 1997a,b), which arises because of the need for firms to anticipate future demand and thus produce forecasts to base their stock control decisions. As the retailer first encounters end-customers demand, it is translated into retailer order. Since members of upper echelons of the supply chain typically are not privy to end customers demand information, they deal with the incoming orders from their downstream partners, which distorts the end-customer demand and thus impacts the quality of upstream forecasts and can cause order amplification. Improvements in the forecasting process can lead to lower costs, depending on the cost structure of the subject organisation (Fildes and Kingsman, 2011), as they target the demand uncertainty. While the increase in the demand variability due to the Bullwhip Effect is connected with more uncertain forecasts at the different stages of the supply chain, it is demand uncertainty, not variability that leads to higher costs; nonetheless these two terms have been conflated often.

The difference between uncertainty and variability has repeatedly been highlighted in the literature (Aviv, 2001, 2002, 2007; Chen and Lee, 2009). Demand variance has been used in many cases as a proxy for demand uncertainty; this is correct only under specific conditions.

Uncertainty relates to the estimation of demand, the magnitude of which relates to the quality of the specified forecasting model. Linking this to the terminology introduced in Chapter 2, Forecast Uncertainty in general refers to Model Uncertainty (MU), while Sample Size Uncertainty (SSU) or Parameter Uncertainty (PU) refer to demand uncertainty. We group these two uncertainties in this category, as in practice it is difficult to disentangle the effect of PU from SSU, and this point is repeated later in this chapter. In the absence of any forecasting mis-specifications, it becomes the intrinsic uncertainty that cannot be accounted for by any model. On the other hand, the variability of demand, which features in the definition of the Bullwhip, measures the spread of observations around their mean. Even when the demand process exhibits a fully predictable deterministic oscillatory pattern, then its variance will be greater than zero. However, since this behaviour is predictable, it will not lead to higher costs as it is fully accounted in the expected demand. Thus, it is the propagation of demand uncertainty, not variability, that is of interest, as it affects the ordering and stock control decisions caused by the Bullwhip Effect.

Traditionally, the Bullwhip effect is detected by the use of a variability-based measure, the Bullwhip Ratio, which compares the variance of upstream demand to that of downstream demand. This does not address forecast uncertainty and its related costs. It only reveals whether upstream and downstream demand differ in variability, but does not diagnose which of the causes of the Bullwhip are responsible for this amplification, nor does it indicate how they contribute to it. Since the ratio does not cover all the areas affected by the order variance amplification, this has lead researchers to propose several ways to either improve it or complement it with other measures by targeting various aspects affected by the Bullwhip (for e.g., Disney and Towill, 2002; Cannella et al., 2013; Chen and Lee, 2009, 2012; Bray and Mendelson, 2015; Trapero and Pedregal, 2016).

This research contributes to the existing literature by studying the impact of demand and forecasting-related uncertainty in the context of the Bullwhip Effect. It stems from the necessity to better understand the impact of forecast uncertainty on upstream stock-control costs, and how it grows along the supply chain. This propagation happens in parallel with the Bullwhip, since the upstream suppliers' forecast errors are expected to be more variable than the retailers', due to the distortion of demand information at different levels of the supply chain, and amount to higher inventory costs for each of the members involved. Examining the behaviour and progression of forecast uncertainty at each node of the supply reveals a link with its associated inventory costs, and how reductions in the latter can be achieved by improving the forecasting process. To this end, we suggest a new metric: the Ratio of Forecast Uncertainties (RFU), consisting of the ratio of conditional standard deviations of the forecast error of the studied member to the retailer's. In order to demonstrate the usefulness of this measure, a supply chain simulation is deployed to determine its relevance in assessing the inventory costs incurred at the manufacturer level, due to the Bullwhip Effect. Our results indicate that the RFU is a good indicator of inventory costs and holds more explanatory power than the conventional Bullwhip Ratio.

The rest of this chapter is organised as follows: Section 3.2 reviews the various studies that have dealt with the demand signal processing aspect of the Bullwhip Effect. Section 3.3 addresses the issue of measuring the Bullwhip Effect, explaining the traditional measurement and highlighting the caveats linked with it. It then proceeds by clarifying the distinction between the concepts of demand uncertainty and variability, to better understand which of them drives cost and should be of interest. Section 3.4 discusses the proposed RFU measure and its interpretation, and justifies the rationale behind its use. Finally, section 3.5 lists the research objectives, and describes the design of the simulation and its findings, accompanied by their managerial implications.

3.2 Modeling the Demand Aspect of the Bullwhip Effect

Ever since the paper by Lee et al. (1997a), four sources have been commonly argued for the Bullwhip Effect: demand forecast updating, order-batching, shortage gaming and price fluctuations (e.g. Lee et al., 1997a). Further causes include the lead-time, the replenishment policy (Dejonckheere et al., 2003; Disney and Towill, 2003a; Jakšič and Rusjan, 2008) and behavioral causes (Croson and Donohue, 2006). Of particular importance is the demand forecast updating, itself intrinsically linked with the definition of the Bullwhip, as forecasts form the basis of ordering decisions. The demand aspect of the Bullwhip Effect can be split in two causes, as suggested by Gilbert (2005): "Type I", where the Bullwhip is present due to the distortion of the demand signal, and "Type II", induced from incorrectly identifying the demand pattern.

3.2.1 Type-I studies

In Type I studies, the underlying demand is assumed to be known a priori, and the relationship between the Bullwhip Effect and other factors, such as the effect of demand correlation or lead-time is examined. In this case, the amplification of variability is not due to incorrectly identifying the DGP (Data Generating Process). To this end, a common assumption is that it is possible to have unbiased and minimum forecast error variance predictions. This is helpful, as it allows in some cases for analytical tractability in deriving closed-forms solutions. The parameters of the known demand process are estimated by employing the Minimum Mean Squared Error (MMSE) estimator. Under this quadratic loss function, the optimal forecast for demand is the conditional mean of its distribution (Gneiting, 2011). It should be noted that using the common Minimum Mean Squared Error criterion ensures these properties for one-step ahead in-sample predictions and not multiple steps ahead (Weiss and Andersen, 1984; Chevillon, 2007; McElroy, 2015), and the stock control implications for this thus remain an open question. Kourentzes et al. (2020) compared the inventory performance of several procedures for generating forecasts, and found that forecasts optimised on the basis of the MSE did not return the optimal performance in terms of coverage rate and inventory trade-off curves. This can be explained as the MMSE only guarantees optimality for in-sample data (where the model and parameters were estimated), but not for out-of-sample data (future unknown data, Makridakis and Winkler, 1989; Barrow and Kourentzes, 2016).

The ARIMA family of models (Box et al., 2015) is the most frequently employed to describe the demand process. Research studies have used: AR(1) (Lee et al., 1997a), constant demand with i.i.d. errors (Chatfield et al., 2004), MA(1) (Ali et al., 2012), AR(p), where p denotes the order of autoregression, (Luong and Phien, 2007), ARMA(1,1) (Gaalman and Disney, 2009), ARMA(2,2) (Gaalman and Disney, 2006), ARIMA(0,1,1), that is equivalent to Simple Exponential Smoothing (Graves, 1999; Babai et al., 2013) and seasonal ARIMA models (Nagaraja et al., 2015; Cho and Lee, 2012). Li et al. (2005) showed that there exists a transition between the Bullwhip and Anti-Bullwhip Effect (defined as the upstream smoothing of order variability), based on the parameters of the ARIMA process. It has also been shown that the resulting orders from these processes will belong to the ARIMA class (Zhang, 2004a; Gilbert, 2005) under certain assumptions. Aviv (2003) provides a general inventory framework based on linear state space models, within which ARIMA is encompassed (Hyndman et al., 2008).

The Martingale Model of Forecast Evolution (MMFE) framework developed by Hausman (1969) and Heath and Jackson (1994) has also received considerable attention in Bullwhip studies. Under this system, it is assumed that demand D follows a martingale, $\mathbb{E}(D_{t+h|t}) = D_t$; and is modeled as the sum of a level term and incremental signals about future demand acquired at each period, themselves assumed to be independent and normally distributed. Apart from the demand structure, the effect of other demand factors have also been investigated in a Bullwhip context, such as the relationship with price (Zhang and Burke, 2011), the effect of substitute products (Duan et al., 2015), market competition (Ma and Ma, 2017), multiple retailers (Sucky, 2009), stochastic lead times (Kim et al., 2006; Reiner and Fichtinger, 2009; Michna et al., 2018), or changes in the product life (Nepal et al., 2012). All these studies reflect the various considerations and difficulties in modeling demand for the Bullwhip Effect, since several factors can impact the upstream amplification of demand variability, either separately or jointly.

3.2.2 Type-II studies

The assumption made in Type I studies allows to study the behaviour of the Bullwhip Effect and the contribution of different variables to it, undeterred by the impact of mis-specifying the forecasting model. However, in most forecasting tasks the true demand is unknown to the modeller, and hence an additional source of error occurs due to incorrectly identifying the underlying DGP (Chatfield, 1995, 1996). Under Type II Bullwhip papers, the demand process is assumed to be incorrectly identified by the forecasting model, a prevalent case in practice, and thus the effect of this mis-specification weighs in additionally in amplifying the customer demand's variance upstream. At the retailer level, this would result in inadequate forecasts, which will be reflected in higher safety stocks and inventory costs (Badinelli, 1990). At higher levels in the supply chain, it is expected that this uncertainty will be further exacerbated as it impacts the retailer orders, which the upstream member will encounter instead of demand. Hosoda and Disney (2009) argue that this could however prove to be beneficial to the manufacturers costs under certain conditions, as it could lead to under-reactions from the upstream member which could offset the magnification of forecast uncertainty.

The works by Chen et al. (2000a,b) examined respectively the impact of using Moving Average and Exponential Smoothing forecasts on order variability when demand follows an AR(1) process. Zhang (2004b) contrasted the Bullwhip produced under an MMSE forecasting method with the aforementioned ones for an AR(1) demand, and found the former to produce the lowest Bullwhip Effect, highlighting the link between better forecasting and lower order variance. Kim and Ryan (2003) quantified the effects of these forecasts in this context on inventory costs, while Sadeghi (2015) extended the study by Zhang (2004b) to a two-product supply chain. These studies confirm that mis-identifying the demand model further exacerbates the already existing Type I Bullwhip Effect, and therefore more accurate forecasts would be beneficial. Nagaraja and McElroy (2018) separate Type II errors into choice of multivariate or univariate forecasting model and choice of forecasting method. A further distinction within the Type II causes should be drawn, differentiating between the case of forecasting parameter uncertainty and forecasting model uncertainty, the latter being currently represented by the Type II papers. The former denotes the instances where the demand form is known but the parameters have to be estimated. This was studied in Pastore et al. (2019b), which found that estimating the autocorrelation parameter for an AR(1) instead of assuming it known a priori led to higher estimates of the Bullwhip.

3.3 Measuring the Bullwhip Effect

3.3.1 Bullwhip Ratio

At the heart of many studies on the Bullwhip Effect lies a central topic: its measurement. The most commonly used metric is the ratio of variances of upstream demand (or orders) to downstream demand (Chen et al., 2000a), denoted from hereon as the Bullwhip Ratio (BWR). The BWR is defined as:

$$\text{BWR} = \frac{\text{Var}(D_t^U)}{\text{Var}(D_t)} \tag{3.1}$$

where D_t^U is the upstream demand, and D_t the customer demand. Since the orders placed by a downstream party will serve as the upstream demand, and assuming that no managerial adjustments or smoothing decisions are made to the orders at that level, the ratio can also be written as:

$$\text{BWR} = \frac{\text{Var}(O_t)}{\text{Var}(D_t)} \tag{3.2}$$

where D_t and O_t denote the customer demand and the orders placed by the firm under study, which is the upstream demand. The BWR can be interpreted as a ratio of variance of the demands in the supply chain, or as a ratio of variance of orders to demand, as in most studies. The Bullwhip exists when BWR > 1, i.e. the variance of upstream demand or orders placed exceeds that of downstream demand, and experiences the Anti-Bullwhip Effect if BWR < 1. This ratio coincides with the noise bandwidth under linear inventory systems (Disney and Towill, 2003b; Wang and Disney, 2016). Other variants have been used in the literature, such as the difference of variances or the ratio of coefficient of variations (Fransoo and Wouters, 2000). Taking the difference of variances corresponds to the additive formulation of the proportional expression in the standard definition. As for the ratio of coefficients of variation, it simplifies to the BWR when both demands have the same mean. However, if potential explanatory variables are omitted from the upstream member's forecasting model, then the estimated model coefficients will be biased (Clarke, 2005), which will affect both the mean and variance of the upstream orders, and this will be reflected in the ratios differing from each other.

The Bullwhip Ratio can represent two aspects, depending on the context and data used: information flow and material flow (Chen and Lee, 2012). Information flow refers to the BWR constructed using demand and orders information, which is the norm in theoretical studies of the Bullwhip, while material flow corresponds to the measure built with sales and shipments data instead of orders and demand. Information flow denotes the transparency of how information translates into orders, and analogously to the Type I - Type II dichotomy, it can be further split into information flow with no forecast distortions (Type I papers), and information flow with forecast distortions (Type II papers). The material flow is encountered in empirical studies such as those that examine whether the Bullwhip is present (Cachon et al., 2007; Bray and Mendelson, 2012, 2015) or whether information sharing among supply chain partners can achieve reductions in the Bullwhip (Trapero et al., 2012; Cui et al., 2015). In these studies, as demand is not observed, it is instead proxied by sales.

Similarly, the orders placed may differ from received shipments, due to upstream capacity constraints or potential production or supply shortages (Chen and Lee, 2012). Furthermore, other factors can affect the final orders such as physical inventory depletion, for instance due to damage, or incentives from decision makers that may lead to a discrepancy between the orders made and the shipment received, which will be reflected in the material flow. The difference between these two measures, and the conditions under which one measure may overestimate the other is investigated in Chen et al. (2017). They advocate that both ratios should be calculated for decisions purposes, as each links to separate costs in the supply chain.

3.3.2 Other measures

While the BWR is the most frequently employed measure in studying the Bullwhip Effect, its scope is limited to determining whether orders are more variable than demand. As a result, it is insufficient on its own. Other measures have surfaced in the literature, some aimed at modifying or refining the current ratio, while others suggesting new metrics to be used alongside the BWR. When replacing orders with production levels, the resulting measure is used to detect the presence of Production Smoothing. Bray and Mendelson (2015) explain that despite sharing the same logic, Production Smoothing and the Bullwhip Effect are two separate supply chain phenomena that can co-exist: the former is the result of production costs which drives production stability, while the latter is the amplification of variance of orders upstream. As a result, they devise a separate metric for Production Smoothing.

The standard form of the BWR assumes that orders and demand are both homoscedastic processes. Motivated by promotions in a retailing context, Trapero and Pedregal (2016) reject it and propose a time-varying bullwhip ratio, which is able to capture the heteroscedastic nature of demand variability. Others researchers have established complementary metrics, with the most common being the Net Stock Amplification Ratio (Disney and Towill, 2002). This measure tracks the amplification of the inventory variance, thus offering a wider perspective on the impact of the Bullwhip Effect. Cannella et al. (2013) propose a measurement system composed of several KPIs (including the BWR and Net Stock Amplification), each targeting a specific aspect of the performance of the supply chain, allowing a holistic assessment of the processes affected by the Bullwhip.

3.3.3 Measurement-related issues

Given that the BWR rests on several assumptions, numerous factors can distort its measurement. Nielsen (2013) studied the robustness of the ratio to the assumptions of normality and mutual independence of orders and demand, and found its performance to deteriorate under small sample sizes. Nagaraja and McElroy (2018) showed that ignoring the cross-correlation between product demands and relying instead on single-product values of the Bullwhip resulted in higher estimates, thus advocating the use of multivariate models to study the phenomenon. Chen and Lee (2012) identified four causes that can impact it: demand seasonality, batch-ordering, the upstream capacity and the level of aggregation. Seasonality is problematic for the BWR as it introduces another source of variability. Due to the imposition of batch sizes, batch-ordering inflates the measure, as the order quantities are rounded upwards. On the other hand, upstream finite capacity dampens the BWR, as it imposes a bound on order quantities.

Unlike other sources of measurement distortion, aggregation pertains to the modeling aspect and the level of granularity of the data, which in turn impacts the measurement of the Bullwhip. It can occur either on a cross-sectional (across products or firms) or on a temporal scale. Rostami-Tabar et al. (2019) established that non-overlapping temporal aggregation leads to a reduction of the Bullwhip Effect for an ARMA(1,1). Fransoo and Wouters (2000) showed that for the same dataset, different values for the BWR can be obtained, based on the level and order of aggregation. Jin et al. (2015a,b) examined the impact of aggregation on the Bullwhip Effect using empirical data, and concluded that aggregation can conceal the BWR. These results are intuitive, given that aggregation is a filter (moving average), and as such weakens high frequency components of the demand series, resulting in smoother signals (Kourentzes et al., 2014). Chen and Lee (2012) proved under certain conditions that as the lead time increases for an ARMA(1,1) process, the value for the Bullwhip ratio converges to one, thus masking the phenomenon. The intuition of this finding is that the uncertainty of the demand process increases for larger horizons, eventually concealing the relative differences in demand variability.

Even though some elements may influence the Bullwhip measurement, the measure itself suffers from a fundamental drawback: its use of the variance of demand. Indeed, due to the existence of lead times, the supply chain member's decisions are based on estimates of demand aggregated at the lead time; and while the variance of orders acknowledges lead time uncertainty, the variance of demand does not. But more importantly, the variance measurement is only meaningful for stationary processes. Many real life time series are not stationary, exhibiting trends and/or seasonal patterns. Additionally, even when the demand series satisfies the condition of stationarity for a certain period, the model can not be expected to remain constant over its entire life span.

For a non-stationary demand process D, the variance is heterogeneous over time ($\mathbb{E}(D_t - \mathbb{E}(D_t))^2 \neq \mathbb{E}(D_{t+h} - \mathbb{E}(D_{t+h}))^2$) and is no longer a meaningful statistic. Even when the mean of the process is constant, it may be heteroscedastic, that is $\sigma_t \neq \sigma_{t+k}$ (see for e.g. Trapero and Pedregal, 2016). In this case the variance estimation that is used in the BWR is misleading. Therefore, these elements must be accounted for when measuring the Bullwhip. One alternative is to process the demand signal so that it is de-trended and/or de-seasonalised. In the presence of trends, an approach similar to the Box-Jenkins methodology can be applied, where a unit root test first identifies whether the data is non-stationary, and then the data is differenced to render it stationary (Bray and Mendelson, 2012; Nielsen, 2013; Wang and Disney, 2016). However, unit root tests, similarly to any test, have limitations to their application depending on the data, as well as varying statistical power. This restricts their general applicability. All considered, the ratio of variances of the differenced series is not the same as the BWR, obfuscating the measurement of the Bullwhip Effect. Another related problem is that removing trend and seasonal components assumes knowledge whether these are deterministic or stochastic, which is difficult to discern with limited sample size (Ghysels and Osborn, 2001). Following an inappropriate decomposition influences further the demand variance by introducing biases, and therefore the BWR.

In addition to the regular time series components, demand often exhibits outliers, for instance due to promotions, and other irregularities. These distort the variance estimation further and need to be treated accordingly (Trapero and Pedregal, 2016). In such cases more robust statistics for dispersion, such as the Median Absolute Deviation, are expected to better capture the baseline demand behaviour, but in this case the irregular periods will be largely under-represented. All these issues, connected with the estimation of demand variability, are all too common in practice, severely limiting the relevance of the BWR for industry. The underlying modelling question is the need for a clear separation between variability and uncertainty, with the latter being more critical for supply chains, as we argue in the following sections.

3.3.4 Variability versus Uncertainty

Uncertainty usually manifests as an increase in unexplained variability of a process. This has resulted in the two terms wrongly being used interchangeably in the literature, with important implications for the measurement of the Bullwhip Effect. There are multiple sources of uncertainty, which can be broadly split into those originating from the DGP and those originating from modelling. Assuming the process possesses some stochastic component, this will introduce some inherent uncertainty. This aspect of uncertainty underlines its connection with the generation of predictions. The stochastic elements of the process have been realised and observed in the past, but remain yet unknown for the future. On the other hand, even with a fully deterministic DGP, it is possible to introduce uncertainty due to imperfect identification and modelling of the target process. Figure 3.1 exemplifies the latter. Given a fully deterministic DGP, a sine wave, we consider four different modelling scenarios. In all cases the variability is the same, as it depends on the mean of the sine wave. Furthermore, as there are no stochastic terms in the DGP, any uncertainty originates from our modelling choices. In scenario (i) the correct model and parameters are used, resulting in zero uncertainty. In scenario (ii) a perfect approximation is used, i.e. a model that can generate sine wave like predictions, but itself is not a sine wave. Again, there is no uncertainty. In a supply chain context, both of these cases can be forecasted perfectly and the demand is covered fully with no need for any safety stock or other risk mitigating actions. Scenario (iii) is quite common in practice, where an adequate approximation is fitted to the observed data, but the

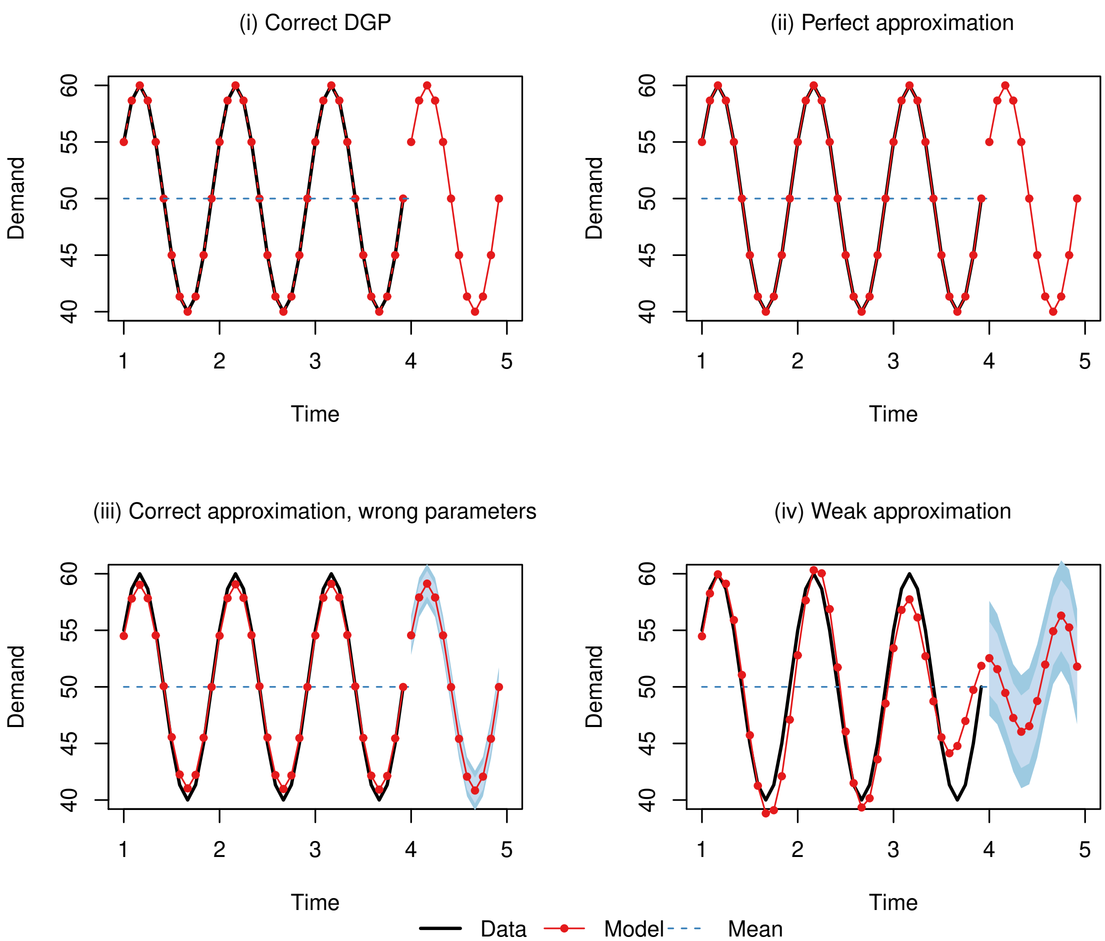

Figure 3.1: Example scenarios contrasting demand variability and uncertainty. All scenarios (i)–(iv) have the same variability. Given a fully deterministic DGP there is no inherent stochasticity, and this is reflected in scenarios (i) and (ii), where either the correct DGP or a perfect approximation is used and there is no uncertainty. In scenario (iii) the approximation is capable of perfectly capturing the DGP, but the parameters are misestimated, resulting in some uncertainty. In scenario (iv) the approximation is weak, resulting in increased uncertainty. In no scenario is the uncertainty connected with the variability.

parameters are imperfect. In this case, uncertainty starts to become relevant, denoted by the shaded area in the subplot. In scenario (iv) the approximation is weak, analogous to the case of using the wrong forecasting model for the given demand signal. Now the uncertainty is much higher. This example demonstrates that demand variability and uncertainty are not connected. When we include stochastic elements in the DGP, this separation may become unclear, as these become part of both variability and uncertainty calculations, yet the illustrated disconnect remains true. Under special conditions the two quantities may appear closely connected (for instance when the demand is i.i.d.), and can result in the misconception that one is a proxy of the other.

In a supply chain context, uncertainty is primarily relevant to predicting the future demand, and as such it is captured by forecasting error metrics, such as the Mean Squared Error (MSE). This highlights a helpful clarification for the nature of demand uncertainty, in that it coincides with the forecast error variability. In fact, for unbiased forecast errors, this becomes the variance of the forecast errors (Saoud et al., 2018). On the other hand, the variance of demand is just a statistical measure of the dispersion of the data points with respect to their mean and therefore unsuitable to capture uncertainty (Fleischhacker and Fok, 2015).

In a more general aspect, consider a demand process $D = (D_t)$, with t being a time index, as a general function of its own lags and explanatory variables X_t, or:

$$D_t = f(X_t, \Omega, \varepsilon_t) \tag{3.3}$$

Here, Ω denotes the model parameters, and ε_t the model innovation or error term, such that $\varepsilon_t \sim N(0, \sigma^2)$. A given set of produced forecasts, $\hat{D}_t$, can be expressed as: $\hat{D}_t = g(\cdot, \widehat{\Omega})$. The variability of demand refers to the unconditional variance of the process D, $\text{Var}(D)$. Aviv (2001) defines uncertainty as the long run conditional variance of demand given a specific forecasting process, which is just the variance of the errors from the forecasting method. This is captured by its MSE, itself conditional on the data available up until time t where it is measured (this definition does not distinguish between in-sample uncertainty and out-of-sample uncertainty, with the latter expected to be higher than the former). The out-of-sample MSE at horizon h is defined as: $\text{MSE}(\hat{D}_{t+h|t}) = \mathbb{E}\left[(D_{t+h} - \hat{D}_{t+h|t})^2\right]$. Other forecasting error metrics can be employed instead of the MSE to measure the level of uncertainty; however the latter denotes the conditional variance at time t of the forecast errors, and is thus selected for analogy to the Bullwhip measurement.

The uncertainty in forecasting demand can be decomposed into two sources: (i) the uncertainty due to estimating the forecasting model, which exists if the functional form underlying D_t and its parameters are incorrectly identified $\left(g(\cdot,\hat{\Omega}) \neq f(\cdot,\Omega)\right)$, as is the prevalent case in real life, (ii) the uncertainty due to estimating the model parameters, which occurs when the structural form of D_t, $f(\cdot)$, is correctly determined, but the parameters Ω have to be estimated, i.e. $\hat{D} = f(\cdot,\widehat{\Omega})$. Each of these uncertainties impacts the overall inventory performance, as they each result in higher inventory costs and higher customer service levels deviations from their intended target (Saoud et al., 2018). More generally, under forecast model uncertainty, the measured error e_t is:

$$e_t = D_t - \hat{D}_t = f(\cdot,\Omega) - g(\cdot,\widehat{\Omega}) + \varepsilon_t \tag{3.4}$$

Consequently, forecast uncertainty can be decomposed into two components: the forecasting model uncertainty and demand uncertainty (Fildes and Kingsman, 2011). The former, $f(\cdot,\Omega) - g(\cdot,\widehat{\Omega})$, represents the deviation from optimality, while the latter represents the demand uncertainty. The model uncertainty can be further expanded to account for the effect of parameter uncertainty, such that:

$$e_t = f(\cdot,\Omega) - f(\cdot,\widehat{\Omega}) + f(\cdot,\widehat{\Omega}) - g(\cdot,\widehat{\Omega}) + \varepsilon_t \tag{3.5}$$

Hence, three components compose the forecast uncertainty: (a) the deviation $f(\cdot,\Omega) - f(\cdot,\widehat{\Omega})$ corresponding to the effect of estimating the model parameters, (b) $f(\cdot,\widehat{\Omega}) - g(\cdot,\widehat{\Omega})$, the additional uncertainty incurred by mis-specifying the true model, and (c) the irreducible demand uncertainty, ε_t. Therefore, adhering to these definitions, forecast uncertainty comprises demand uncertainty, and this terminology is adopted throughout the remainder of this thesis.

As demand updating constitutes one of the four original causes of the Bullwhip, it is a result of forecast uncertainty, affecting the subsequent order quantities and stock levels.

Therefore, it is uncertainty, not variability, that is the cost driver. To illustrate the cost impact of uncertainty, Aviv (2001) provided an example of a demand which follows a deterministic pattern, but can be fully predictable by the upstream member. As incoming demand can be perfectly foreseen, there exists no demand or forecast uncertainty and hence, no additional inventory costs associated with it. However, demand points can still oscillate around the mean, so the variability of demand would nonetheless be present, even in the absence of uncertainty. This echoes the example given in Figure 3.1. While it has been shown under a specific set of assumptions that the BWR is linked to the upstream costs (Chen and Lee, 2012), this relationship does not hold in most cases, as the BWR should not necessarily be linked to inventory costs, following the discussion on variability versus uncertainty. In addition, dampening the variability of orders does not necessarily result in lower costs, since it might not lower the uncertainty level (Chen and Lee, 2009). For instance, Chen and Samroengraja (2004) assess whether replenishment policies aimed at mitigating the Bullwhip Effect are the most effective, and conclude it to not be always true, since lowering the variability of orders does not imply a decrease in the uncertainty level, and thus costs are not expected to always decrease as a result of this strategy.

This indicates that the focus of solutions aimed at reducing the cost should be shifted to reducing the propagation of uncertainty along the supply chain, since demand forecasting is at the forefront of many inventory and orders decisions. Some authors have advocated employing different metrics alongside the BWR. Aviv (2001) suggest pairing it with demand uncertainty metrics, while Bray and Mendelson (2012) provide a Bullwhip estimator, based on the conditional variance of MMFE forecast errors at different lead times. The need for an uncertainty-related metric for the Bullwhip prompted this research. In this chapter, rather than refining the BWR due to its shortcomings, a new method for capturing the amplification of forecast uncertainty is proposed, based on forecast error measures. It addresses the forecasting impact and how its uncertainty evolves over the supply chain in the presence of the Bullwhip.

3.4 Proposed Measure

The new metric proposed in this thesis is based on determining the Ratio of Forecast Uncertainties (RFU) between the upstream tier and the retailer. Given that the first forecast occurs at the retailer, this ratio tracks the progression of forecast uncertainty, represented by the cumulative Root Mean Squared Errors (CumRMSE), which is the sample standard deviation of the observed forecasting errors, aggregated at the horizon under study (Syntetos and Boylan, 2006; Saoud et al., 2018). The ratio thus benchmarks the variability of the upstream's forecast errors to the downstream's, and tracks similarly to the BWR, the evolution of forecast uncertainty.

At any level of the supply chain, given D_t, the demand up until period t, and its corresponding forecasts $\hat{D}_t$, the cumulative RMSE over horizon H is defined as:

$$\text{CumRMSE} = \sqrt{\frac{1}{(N-H+1)} \sum_{t=1}^{N-H+1} \left(\sum_{i=1}^{H} D_{t+i} - \sum_{i=1}^{H} \widehat{D}_{t+i|t} \right)^2} \tag{3.6}$$

where N is the number of observations, H the forecast horizon (defined for a periodic inventory review model as lead time + review period). Unlike the standard definition of the RMSE, (3.6) captures the cumulative uncertainty of forecast errors over the horizon (Syntetos and Boylan, 2006), which is necessary for inventory decisions as it takes into account the correlation of forecast errors at different horizons (Saoud et al., 2018). Subsequently, the RFU between the upstream echelon U and the retailer R is determined as:

$$\text{RFU} = \frac{\text{CumRMSE}^U}{\text{CumRMSE}^R} = \sqrt{\frac{(n-l+1)}{(N-L+1)} \frac{\sum_{t=1}^{N-L+1} \left(\sum_{i=1}^{L} D^U_{t+i} - \sum_{i=1}^{L} \widehat{D}^U_{t+i|t} \right)^2}{\sum_{t=1}^{n-l+1} \left(\sum_{j=1}^{l} D^R_{t+j} - \sum_{j=1}^{l} \widehat{D}^R_{t+j|t} \right)^2}} \tag{3.7}$$

where L and l denote the respective protection intervals (the sum of the lead time and the inventory review period) under consideration for the upstream member and retailer, and N and n represent the respective sample sizes.

By employing the proposed ratio, the propagation of demand uncertainty can be estimated at the desired level of the supply chain. The RFU accounts for the lead times of both supply chain members, which is not the case for the BWR. Similarly to the Bullwhip Ratio, demand uncertainty is increasing in the supply chain for RFU > 1, steady for RFU = 1, and decreasing otherwise. It is expected that for a common lead time, the ratio will be greater than one, provided reasonable forecasts, as the orders placed (which serve as the upstream demand) would be expected to be more volatile than the initial demand, and hence more uncertainty in forecasting it.

The RFU is not restricted to a common forecast horizon (the sum of the lead time and the review period of the ordering policy), as it takes into consideration both horizons under which the supply chain partners are operating, an advantage it holds over the BWR which only features the upstream lead time. For example, if the supply chain under study consists of a retailer with a small horizon and a wholesaler or manufacturer with a greater horizon, the value for the RFU is anticipated to be high. Hence, the value of the horizons and the difference between them can affect the ratio. Equation 3.7 indicates that increasing the manufacturer's horizon, ceteris paribus, will lead to a higher RMSE value in the numerator which will inflate the RFU, assuming a monotonic increase in forecast errors. For the retailer, the impact of changing his horizon is not straightforward. On one hand, increasing his lead time will lower the denominator of the ratio, but this will also manifest itself in the upstream member having a higher forecast uncertainty as a result.

Relative error measures, which consist of the ratio of one metric to another, are common in the forecasting literature. They are simple to interpret, as they assess the performance of one method to a benchmark, and are scale independent (Hyndman and Koehler, 2006). Theil's U statistic (Theil, 1966) compares the RMSE of one-step-ahead forecasts to those made from a Random Walk model, but quite easily the benchmark could be replaced with any model. The Relative Geometric Root Mean Square Error (Relative GRMSE) examines the ratio of the root of the geometric mean of squared errors of forecasts against those of a benchmark model at the relevant forecast horizons (Fildes, 1992). Hyndman and Koehler (2006) showed that this is the same as the Geometric Mean Relative Absolute Error (GMRAE), that is the geometric mean of the ratio of absolute errors of two competing forecasts. Davydenko and Fildes (2013) proposed the Average Relative Mean Absolute Error (AvgRelMAE), that is the geometric mean across items of the ratio of mean absolute errors of two competing forecasts. The main difference between AvgRelMAE and GMRAE is that the former is calculated on already summarised absolute errors, hence mitigating most computational issues, but having somewhat less sensitivity. Kourentzes and Athanasopoulos (2019) addressed the sensitivity issue by suggesting the AvgRelRMSE, in the same spirit of AvgRelMAE. The absolute loss tracks the median of the distribution of errors, while the quadratic loss tracks the mean, being somewhat more sensitive (Gneiting, 2011). The difference between RFU and AvgRelRMSE is that the former uses cumulative RMSE figures over lead time, and therefore is directly connected to the inventory decisions. The quadratic loss in RFU is a natural choice, as it is directly involved in the determination of safety stocks and the resulting orders and inventory levels.

This definition of the ratio of RMSEs should be distinguished from that found in Ouyang and Daganzo (2006, 2008); in the latter, this ratio denotes the ratio of standard deviations of orders to demand, i.e. the square root of the BWR. In this thesis, the RMSE is the conditional standard deviation of the forecast errors, estimated from a finite set of observed demand and forecasts. All sources of forecast uncertainties discussed in the previous section are captured by the MSE. From the Bias-Variance decomposition, it can be seen that the MSE encompasses both bias and variance of the forecasts, and it reduces to the latter under the assumption of unbiased forecasts, capturing the various mis-identification forms of the forecasts. Since forecast bias has been determined to be the main cost driver for inventory decisions (Zhao and Xie, 2002; Sanders and Graman, 2009, 2016; Wan and Sanders, 2017), the MSE is expected to bear a more direct relation with the inventory costs than the variance of demand. Given that safety stocks are calculated based on the RMSE rather than the MSE, the ratio of the former is preferred to the latter. The RMSE is determined at the lead-time, thus aligning the proposed ratio with the decision event horizon (Chatfield, 2000). This measure incorporates both the lead times of the retailer and that of the upstream member. It holds a crucial advantage over the Bullwhip Ratio, being able to handle non-stationary or seasonal demands. When reasonable forecasts are used, the residuals are expected to be stationary, irrespective of the structure of the demand. Therefore, it overcomes one of the key limitations of the BWR.

As the BWR and RFU represent conceptually separate phenomena which are nonetheless inter-twined, comparing the two metrics is non-trivial. To approach this, we focus on the cost implications of the Bullwhip Effect, specifically on inventory costs. While improved forecasting accuracy brings forth lower costs (Zhao and Xie, 2002), the linkage between the two is not a direct one, as decreases in the former are not translated into equal decreases in the latter (Flores et al., 1993; Babai et al., 2013). Improvements in the forecasting process lead to lower values of the BWR as evidenced by the studies of Zhang (2004a) and Wright and Yuan (2008), since more accurate forecasts will entail less forecast uncertainty and hence less variable upstream orders. Chiang et al. (2016) found no link between the Bullwhip and some error metrics in their empirically simulated study of the Bullwhip Effect in the automative industry, yet it should be noted that they measured forecasting accuracy using the Mean Absolute Percentage Error (MAPE) and not the MSE. Despite these findings, we expect that the RFU has a connection with the upstream tier's inventory costs, as it contains information pertaining to both that level and the downstream's forecasting accuracy.

From a managerial perspective, the RFU is relevant for improving supply chain performance, as it keeps track of the forecast uncertainty of both the upper member and the retailer, and can be used in conjunction with other supply chain key performance indicators. It allows monitoring the propagation of forecasting and hence demand uncertainty along the supply chain. It also enables the assessment of the relative gains by the upper echelon in terms of reduction in uncertainty by improving the forecasting process or including other explanatory market signals, such as customer demand or promotional plans. It can thus be used as a tool to assess the potential forecasting benefits of Information Sharing between the two echelons, as the gains from additional information will be reflected in reductions in the forecast uncertainty of the upper member, and the RFU values will be expected to be in the range of 1 for matching lead times, i.e. no upstream propagation of forecast uncertainty.

This metric is actionable, following the definition given by Aviv (2007), as the parties involved can take action upon it by changing the forecasting process at the desired level. In the case of the BWR, measures can be taken to lower the numerator (order variability) such as reducing the lead time and/or order batch size (Lee et al., 1997b), sharing information (Ali et al., 2012), smoothing orders (Balakrishnan et al., 2004), obtaining advanced demand information (Kunnumkal and Topaloglu, 2008), scheduled ordering (Cachon, 1999; Kelle and

Milne, 1999) or postponing orders (Chen and Lee, 2009). However, taking actions to change the denominator in the BWR (the demand variability part) proves to be more difficult, as demand is exogenous to the firm, and reducing its variability would entail modifying the customer's behaviour (Fildes and Kingsman, 2011).

3.5 Simulation

3.5.1 Objectives

We employ a dyadic supply chain simulation to examine the properties of the proposed measure. We focus particularly on the impact of the nature of the underlying demand process, the forecast horizon of the retailer and of the upstream member, as well as other secondary factors. The calculation of RFU involves the cumulative RMSE of the two members of the simulated supply chain, which itself is connected with safety stock calculations and ensuing orders. Therefore, we anticipate a closer connection between RFU and inventory costs, than the BWR. As previously discussed, the different limitations of the BWR manifest themselves in different demand conditions. In our simulation, we consider a variety of demand patterns and lead times to highlight the comparative behaviour of BWR and RFU in favourable and unfavourable conditions.

3.5.2 Experimental Design

To approach the goals set previously, a supply chain simulation was devised. The selection of this methodology over an empirical approach stems from the necessity to exert control over the nature of the demand processes and the supply chain structure. The objectives of this simulation are to: (i) get a better understanding of the distribution of the proposed RFU metric under different experimental settings and (ii) study which of demand uncertainty and variability, represented respectively by the RFU and BWR), is more related to the upstream member's inventory costs.

We model a dyadic supply chain, consisting of demand for a single-item SKU with one retailer and one manufacturer. The retailer observes their demand, and places their orders to the manufacturer. The manufacturer observes the incoming retailer orders and places their orders from an external source.

We consider three monthly demand processes from the ARIMA family:

Table 3.1: Data Generating Processes

p	Process	Functional form
1	AR(1)	$(1-\phi B)D_t = \varepsilon_t.$
2	IMA(1,1)	$(1-B)D_t = (1+\theta B)\varepsilon_t.$
3	$\text{ARIMA}(0,1,1)(0,1,1)_{12}$	$(1-B)(1-B^{12})D_t = (1+\theta B)(1+\Theta B^{12})\varepsilon_t$

where B is the backshift operator. The first DGP corresponds to the AR(1) model featured in many studies, and is a stationary process, where many of the shortcomings of the BWR discussed in section 3.3 will not appear strongly. The other two models are non-stationary, making the calculation of demand variance invalid and therefore weakening the BWR. The first non-stationary generating model considers the case of a stochastic local level, and is equivalent to the Simple Exponential Smoothing model, that is very commonly used in practice (Gardner Jr, 2006; Ord et al., 2017); while the second, further includes stochastic seasonality, matching the well known Airline model (Box et al., 2015).

For all three demand models, the innovation term $\varepsilon_t \overset{i.i.d}{\sim} N(0,\sigma^2)$. The model parameters $\{\phi,\theta,\Theta\}$ are sampled to guarantee stationarity and invertibility (Box et al., 2015). A constant level is added to all three DGPs to ensure non-negative values. We generate 500 observations for each series and partition them into three subsets: the first 200 points make up the training set over which a forecasting model is identified and fitted; the next 200 points serve as a burn-in period used to eliminate any initialisation bias of the inventory simulation; and the last 100 points are used as a test set, on which the evaluation is conducted. The RFU is calculated using equation (3.7), and the BWR using (3.2). For each process, 500 replications are produced, ensuring a representative sampling of the output of the inventory simulations.

Simulating real practice, where the data generating process of the demand is unknown, we opt to identify an appropriate ARIMA for each time series. We follow the modelling methodology by Hyndman and Khandakar (2008) who specify a stepwise search of ARIMA orders using the Akaike Information Criterion corrected for small sample sizes. Therefore, model mis-specification and parameter estimations issues may be present. All forecasts were produced following a rolling origin scheme, with the model and parameters estimated once over the training set and no longer recalibrated (Ord et al., 2017).

We consider three values for the retailer's horizon, h, and the manufacturer's, H, such that $h, H \in \{1,3,5\}$. The target cycle service level for each member of the supply chain is set at: 90%, 95% and 99%, reflecting values used in practice. We introduce order batching scenarios at the retailer's level with batches b of 1, 10 and 20 units, as it has been identified as a source of the Bullwhip Effect as well as contributing to the distortion of the BWR (Chen and Lee, 2012). Hence, final orders O'_t are rounded up from the prescribed orders from the inventory policy O_t as multiples of the batch size b for the retailer, such that $O'_t = b\lceil\frac{O_t}{b}\rceil$. We use {1, 10, 50} for the distribution of the error term of the DGP, s, in order to represent cases with low, medium and high levels of demand variability. The full set of control parameters for the experimental design are listed in the table below.

Variable	Values	Options
Downstream Demand Process	AR(1), IMA(1,1), ARIMA(0,1,1)(0,1,1)$_{12}$	3
Demand Noise Level	1, 10, 50	3
Retailer Horizon	1, 3, 5	3
Manufacturer Horizon	1, 3, 5	3
Retailer Service Level	90%, 95%, 99%	3
Manufacturer Service Level	90%, 95%, 99%	3
Order Batches	1, 10, 20	3

Table 3.2: Experimental Design Control Parameters

For each of the retailer and the manufacturer, a periodic Order-Up-To replenishment policy is implemented, with review period R set at 1, and unmet demand is assumed to be backordered. The Order-Up-To level is calculated as:

$$S = \sum_{i=1}^{H} \widehat{D}_i + k_\alpha \widehat{\sigma}_i^{SS} \tag{3.8}$$

where $\widehat{D}$ is the demand forecast, H the horizon of interest, k_α is a factor from the inverse of the normal distribution associated with the desired cycle service level α, and the standard deviation of the forecast errors over the lead time, $\widehat{\sigma}_{SS}$ is the CumRMSE defined earlier calculated using (3.6) as argued for by Saoud et al. (2018). As the simulation rolls over the available sample, $\widehat{\sigma}_{SS}$ is updated, making use of the newly available forecast errors.

For the performance evaluation, we monitor the inventory levels for the upstream supply chain tier. We consider only inventory related costs, and these are composed of the costs of holding excessive inventory and those of backordering demand. There are no fixed ordering costs in our simulation. Given the difficulty in estimating backorder costs, we calculate them by approximating the cycle service level with the fill rate, similarly to a Newsvendor problem. Given the CSL approximation provided, the backordering cost can be determined by plugging in the CSL, and setting the holding cost to 1. For a given cycle service level, CSL, the following approximation holds:

$$CSL = \frac{c^+}{c^+ + c^-}, \tag{3.9}$$

where c^+ and c^- denote the costs of excessive and backordered demand respectively. The holding cost, c^+, is set as 1 and $c^- = \{1, 19, 99\}$. The simulation was conducted using the R statistical language (R Core Team, 2019).

3.5.3 Results

Given the large number of dimensions in this experiment, a preliminary analysis was conducted first to determine which factors introduce substantial variation in the results. A careful inspection revealed that the noise level of the demand and the retailer service level do not drastically affect the studied measures. Furthermore, while larger order batch size leads to higher RFU and BWR values, as well as higher total costs, the increase in total costs with respect to the other two metrics was found to be proportional for all cases. When order batching is introduced, the BWR increases (as the orders become more variable and this is carried through upstream) and total costs increase. However, when looking closely at the results, and comparing the effect of order batching on either metrics, we notice that for all metrics, the increase happens in a proportional way. Therefore, the results were averaged across service levels, the noise level and the size of the order batches. The remaining dimensions of interest are the demand process type, and the forecast horizon for each of the retailer and the manufacturer. Now that we have identified these variables, we wish to examine how each affects both the RFU and the BWR, as well as the total manufacturer's cost.

The distinction between demand uncertainty and variability becomes evident in Table 3.3, which provides the ratio MSE_{t+h}/Var(Demand), i.e. the ratio of demand uncertainty over conditional variability, for each tier of the supply chain, for the different simulation cases. The ratios show that, not only is there a difference between demand uncertainty and variability, but also that the difference may be considerable. The retailer values do not change for different manufacturer lead times and therefore are omitted to avoid cluttering the table. Considering manufacturer lead time of 1, observe how the ratio changes between demand process. For the stationary case, as the forecast horizon increases forecast uncertainty dominates. For the non-stationary cases, the variability of the predictable demand overwhelms the forecast uncertainty resulting in very low values, particularly for the ARIMA(0,1,1)(0,1,1)$_{12}$. This highlights the main weakness of using demand variability instead of forecast uncertainty, as the easily predictable part of the demand dominates the measurement. As the lead time for the manufacturer increases, the forecast uncertainty increases as well, increasing the value of the ratio. Comparing the retailer's and manufacturer's figures, the variability of the demand of the latter is increased over the former, due to the transformation of the demand by the ordering process of the retailer.

Table 3.3: Ratio of MSE to variance of demand for each echelon in the simulation at all configurations.

Process	Echelon	Manufacturer horizon 1			Manufacturer horizon 3			Manufacturer horizon 5		
		Retailer horizon			Retailer horizon			Retailer horizon		
		1	3	5	1	3	5	1	3	5
AR(1)	Retailer	0.63	4.17	10.26	–	–	–	–	–	–
	Manuf.	0.62	0.88	0.99	4.43	3.57	3.37	9.57	8.11	6.42
IMA(1,1)	Retailer	0.17	1.25	3.94	–	–	–	–	–	–
	Manuf.	0.19	0.55	0.80	1.78	2.51	3.26	5.12	7.27	7.00
ARIMA	Retailer	0.01	0.02	0.05	–	–	–	–	–	–
(0,1,1)(0,1,1)$_{12}$	Manuf.	0.01	0.04	0.06	0.05	0.22	0.40	0.11	0.41	0.91

Figure 3.2 illustrates the RFU distributions for different processes and horizons, for the retailer and the manufacturer. The figure provides violin plots, to better represent any asymmetries and extremes of the distributions. These are grouped in three sets, one for each of manufacturer's horizons, and within each set there are three subsets, one for each of the retailer's horizons. The differently colour coded violin plots refer to different demand processes. The greyed area corresponds to the case of RFU < 1.

In all cases, as the process becomes more complex, and crucially from stationary to non-

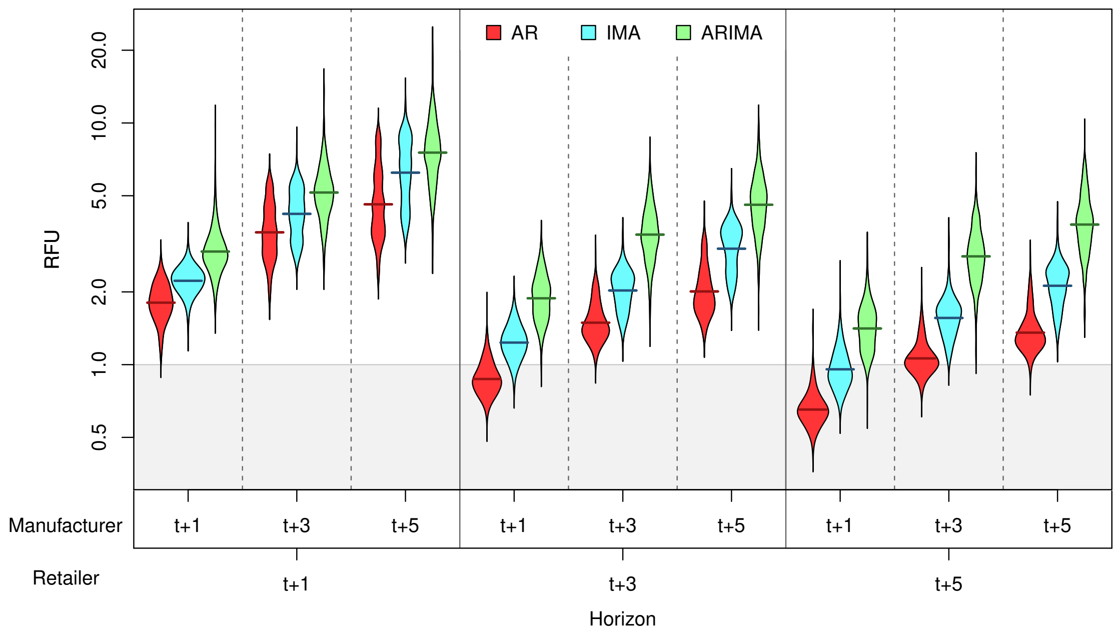

Figure 3.2: Violin Plots of the RFU per demand process, colour-grouped by process

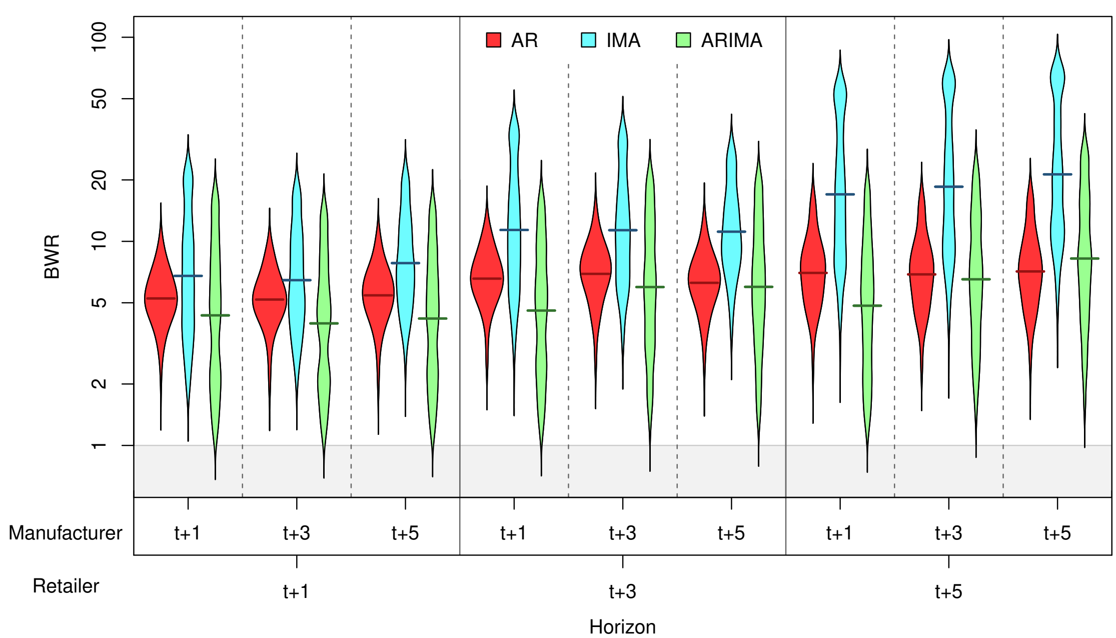

Figure 3.3: Violin Plots of the BWR per demand process, colour-grouped by process

stationary, the location of the RFU distribution shifts to higher values. Furthermore, distributions become wider. In terms of RFU, that means that although the forecast errors for the retailer increase, due to the more complex demand process, the demand observed and modelled by the manufacturer is even more difficult to forecast, since the initial demand is transformed by the ordering process of the retailer. As the lead time of the retailer increases, so within each set of distributions for a given manufacturer horizon, the distortion of the initial demand signal increases, making forecasting for the manufacturer more difficult, reflected in higher RFU values. On the contrary, as the manufacturer lead time increases, forecasting becomes more difficult, offsetting the modelling difficulties due to any distortions from the ordering process of the retailer, effectively lowering RFU. The violin plots indicate that the RFU describes well the relative forecasting difficulty, as captured by the forecast uncertainty of the retailer and the manufacturer in the different cases. It is also interesting to observe when RFU is lower then one, suggesting that the relative forecast uncertainty decreases. This occurs for the majority of the cases of the stationary demand and manufacturer forecast horizon of 1, when the retailer lead time is longer than a single period. In these cases the forecast errors of the retailer remain small, while the longer horizon for the manufacturer introduces more uncertainty, lowering the RFU below 1. This is stronger for the stationary case, although not absent in the non-stationary ones.

We contrast Figure 3.2 with Figure 3.3 that provides the BWR presented in a similar way. Observe that the differences in lead times, for either the retailer or the manufacturer cause small only changes. As expected, the BWR is unable to capture the forecast uncertainty and reflect the implied forecast risk for either member of the supply chain. Furthermore, the differences between the distributions of the BWR of the various processes are small as well, even though the associated forecastability is substantially different, as demonstrated in Table 3.3.

The differences in how RFU and BWR represent the supply chain simulations, highlight one of the goals of this research: to determine the relevance of the BWR and the RFU as indicators of total inventory costs. We model the total inventory costs as:

$$C_i = a_1 X_i + \underbrace{\sum_{j=1}^{2} a_{1+j} L_{i,j} + X_i \sum_{j=1}^{2} a_{3+j} L_{i,j}}_{\text{Manufacturer lead time}} + \underbrace{\sum_{j=1}^{2} a_{5+j} R_{i,j} + X_i \sum_{j=1}^{2} a_{7+j} R_{i,j}}_{\text{Retailer lead time}} + \underbrace{\sum_{j=1}^{3} a_{9+j} P_{i,j} + X_i \sum_{j=1}^{2} a_{12+j} P_{i,j+1}}_{\text{Demand process}} + \eta_i, \quad (3.10)$$

where C_i denotes total inventory costs at the manufacturer, which is the sum of holding costs and backorder costs, for observation i of the simulation, $L_{i,j}$ is a set of binary dummies encoding the manufacturer's lead time ($L_{i,1}$ corresponds to $H = 3$ and $L_{i,2}$ corresponds to $H = 5$). Similarly, $R_{i,j}$ encodes the lead time for the retailer. The set of binary dummies $P_{i,j}$ encode the demand process, where j corresponds to the cases of p in Table 3.1 and η_t are i.i.d. errors. Finally, X_i can either be RFU resulting in model $M1$, or BWR for $M2$. Observe that we retain the case for all three j values of $P_{i,j}$, as an intercept is absent in the model. This is the case as there are no fixed costs in our simulation. Equation (3.10) includes interaction terms between X_i and the dummy variables to account for non-linearities. For the same reason, we re-code the lead times into dummy variables. The model form for both M1 and M2 were validated using the Akaike Information Criterion, which indicated that two-way interaction terms were sufficient.

Table 3.4 displays the coefficient estimates and their associated p-values for both models, together with their coefficient of determination R^2. The latter indicates that M1, which relies on RFU, explains better the total inventory cost than M2 that uses the BWR. Note that for M1 all terms, apart from the interaction between RFU and M_2, are significant. This is not the case for M2, where most of the terms that include X, i.e. the BWR, are insignificant or weakly significant. For M2 the variability of the total inventory cost is mostly explained by the dummy variables for the lead time of the manufacturer and the demand process.

Although the regression summary in Table 3.4 is indicative that RFU explains the total inventory cost, C, better than using BWR, it is difficult to understand the nature of the connection with C. In Table 3.5 we provide the net effect of X inclusive of any interaction terms, so each lead time and process case in our simulation. Since the experimental settings vari-

Table 3.4: Regression summary for M1 and M2

Variable		M1 (RFU)		M2 (BWR)	
X	a_1	-204012.82	***	274.94	
$H = 3$	a_2	-109663	***	76988.88	***
$H = 5$	a_3	-339080.3	***	534984.4	***
$X(H = 3)$	a_4	69218.07	***	-1616.62	*
$X(H = 5)$	a_5	221540.9	***	-3269.7	***
$h = 3$	a_6	-334320.39	***	-28224.2	
$h = 5$	a_7	-471148.17	***	-8820.89	
$X(h = 3)$	a_8	130735.56	***	1730.95	
$X(h = 5)$	a_9	261414.55	***	2182.77	
M_1 (AR)	a_{10}	441017.14	***	-168010	***
M_2 (IMA)	a_{11}	422229.85	***	-133077	***
M_3 (ARIMA)	a_{12}	142545.94	**	414921.7	***
XM_2	a_{13}	8563.85		-222.78	
XM_3	a_{14}	191963.84	***	-5190.37	**
R^2		0.416		0.312	

***p-value < 0.001, **p-value < 0.01, *p-value < 0.05

ables in our regressions are coded as dummy variables, and given the existence of interaction terms, the net effects allows us to determine the total effect of a given variable with respect to the independent variable, which is total upstream inventory costs. The provided values are the sum of the coefficients from Table 3.4 for each scenario, excluding the additive effects of the lead time dummies without any interactions (coefficients a_2, a_3, a_6, a_7). The demand process dummies are retained, as they accumulate any remaining effects in the absence of an intercept term.

Observe that the connection of RFU is always positive. As the ratio of forecast uncertainty increases, so do the total inventory costs. This signifies that as forecasting becomes worse, i.e. the uncertainty increases, then it is necessary to accept more inventory costs to account for the increased risk. This follows from the arguments that we based the proposed metric. On the other hand, when we look at the net effects for the BWR we cannot get a consistent view. For the highly variable ARIMA(0,1,1)(0,1,1)$_{12}$ process, the BWR is positively connected to C (notice that this is mainly driven by a_{12} in Table 3.4), while for both AR(1) and IMA(1,1) the BWR is negatively connected. Furthermore, for the case of RFU as the lead time increases, which requires additional safety stocks and therefore implies additional costs, the net effect increases as expected. This is not the case for the BWR, where the lead time seems to have a minimal influence on the net effect of BWR on C. This mirrors our understanding from

Table 3.5: Net effects of RFU and BWR on total inventory cost

Lead time		Process		
Manuf.	Retailer	AR	IMA	ARIMA
		M1 (RFU)		
$H=1$	$h=1$	237004.32	226780.88	130497
	$h=3$	367739.88	357516.44	261232.5
	$h=5$	498418.87	488195.43	391911.5
$H=3$	$h=1$	306222.39	295998.95	199715
	$h=3$	436957.95	426734.51	330450.6
	$h=5$	567636.94	557413.5	461129.6
$H=5$	$h=1$	458545.22	448321.78	352037.9
	$h=3$	589280.78	579057.34	482773.4
	$h=5$	719959.77	709736.33	613452.4
		M2 (BWR)		
$H=1$	$h=1$	-167734.65	-133024.84	410006.2
	$h=3$	-166003.7	-131293.89	411737.2
	$h=5$	-165551.88	-130842.07	412189
$H=3$	$h=1$	-169351.27	-134641.46	408389.6
	$h=3$	-167620.32	-132910.51	410120.6
	$h=5$	-167168.5	-132458.69	410572.4
$H=5$	$h=1$	-171004.35	-136294.54	406736.5
	$h=3$	-169273.4	-134563.59	408467.5
	$h=5$	-168821.58	-134111.77	408919.3

Figures 3.2 and 3.3.

Although the analysis so far demonstrates the ability of the RFU to better capture the total inventory costs, this does not preclude the usefulness of the BWR for understanding the costs. To investigate this we construct an encompassing test, between models M1 and M2, where we test whether the residuals of a model can be explained by the other, signifying that there is information in the residuals that could be modelled better by the inclusion of BWR or RFU into M1 or M2 respectively (Fang, 2003). The purpose of a forecast encompassing test between two competing models is to determine whether one model contains additional information over the other in predicting the dependent variable shared by both models. In other words, we are assessing whether a model can be improved by using the other independent variable as an additional regressor in predicting total upstream inventory costs.

To this end, we adopt the test proposed by Chong and Hendry (1986), into the following two regressions models:

$$\text{M3}: C_i - \hat{F}_{M_1,i} = \beta \hat{F}_{M_2,i} + \epsilon_{M_1,i}, \tag{3.11}$$

$$\text{M4}: C_i - \hat{F}_{M_2,i} = \gamma \hat{F}_{M_1,i} + \epsilon_{M_2,i}, \tag{3.12}$$

where $\epsilon_{M_1,i}$ and $\epsilon_{M_2,i}$ are normal i.i.d. terms, $\hat{F}_{M_1,i}$ and $\hat{F}_{M_2,i}$ are the predicted values of M1 and M2 respectively. For M3 if β is significant then RFU should be used together with BWR to explain the total inventory costs. In analogy, if γ is significant, this suggests that BWR should not be used on its own and RFU is necessary.

We find β to be equal to 0.01 with a p-value of 0.61, while γ is equal to 0.34, with an associated p-value of less then 0.001. Therefore, we conclude that there is no evidence that BWR adds to explaining the total inventory costs when we use RFU. The same is not true for BWR and there is evidence that RFU is necessary to better explain the costs. This findings is in agreement with the regression summaries provided in Tables 3.4 and 3.5, where BWR was found to be mostly insignificant and its net effect on C was counter-intuitive.

3.6 Conclusion

The Bullwhip Effect has received considerable attention within the field of supply chain management. Given the many deleterious consequences associated with it, accurate quantification is necessary in order to buffer against it. To this end forecasting is central, as producing estimates of the future demand and consequently orders affects how the end-customer demand is translated for the upstream members of the supply chain. In producing accurate forecasts, it is the demand uncertainty, rather than the demand variability that is critical. We stress that even though the amplification of the latter is attributed to the Bullwhip, the former is the cost driver, and should receive more attention in the context of the Bullwhip Effect. This permits us to obtain a better grasp on the impact of forecasting on the Bullwhip and its resulting costs.

We propose a new metric, the Ratio of Forecast Uncertainty, that is built on forecast uncertainty, rather than demand variability. It circumvents some of the drawbacks of the Bullwhip measure, such as its ability to handle non-stationary demand, which is often encountered in real life scenarios. Furthermore, it goes beyond other similar metrics, incorporating relevant forecast lead times, matching the forecasting tasks that the members of supply chain face.

Through an inventory simulation, we explore the properties of the RFU and contrast them with those of the BWR. We find that the complexity of the demand process and the upstream lead-time increase the forecast uncertainty, as seen by higher values of the RFU. In order to assess the practical relevance of the proposed metric, we compare it with the BWR in predicting total upstream inventory costs. Our results suggest that the RFU is linked with the aforementioned costs, and holds more power over its counterpart in explaining inventory costs. This is pronounced especially in the case of non-stationary demands, where the BWR is no longer a meaningful indicator, a fact confirmed by our findings. Furthermore, we use encompassing tests to show that there is no additional benefit arises from using both metrics in understanding costs over the use of the RFU solely; however the converse does not hold. Nonetheless, note that although RFU encompasses BWR, a large part of the inventory costs remain unexplained, suggesting that other complimentary metrics to BWR discussed in the literature remain useful.

From a managerial standpoint, the RFU allows to determine the rate at which the overall forecast uncertainty in the supply chain grows, but also reflects the upstream impact of certain measures aimed at decreasing the overall forecast uncertainty, such as either (or both) members reducing their lead time, or improving the current forecasting methodology. The later may include better forecasting models, supplementing forecasts with explanatory variables, judgemental adjustments and so on. The RFU facilitates connecting the effort put into improving the forecasting process to the inventory costs. In that sense, a fertile area to use RFU is investigating the effects of collaboration and information sharing.

Acknowledgments

All calculations were performed on the High End Computing Facility at Lancaster University, United Kingdom.

Chapter 4

The Forecasting and Inventory Value of Information Sharing under Different Modelling strategies

Demand information sharing has been proposed as an effective solution to counter the Bullwhip Effect in supply chains, as it offers improvements in forecasting accuracy and reductions in inventory. With the availability of the customer demand data, the upstream member in the supply chain can generate forecasts on it directly instead of conventionally using historical orders. While this approach has returned favourable results in most analytical studies, it effectively ignores information that might be contained in the orders history, but absent from the demand data, such as managerial interventions. Some studies have identified alternative multivariate ways to use both orders and sales in the forecasting model, but there is very limited comparison between alternatives. Furthermore, the focus in the literature has been primarily on forecast accuracy. We design a simulation evaluation to compare the information sharing alternatives, on accuracy and inventory costs this paper. Forecast uncertainty is introduced at all echelons, as well as managerial orders adjustments, in order to better reflect practice. Relying solely on point of sales information performs best in all scenarios, while the multivariate approach still adds value over no information sharing.

4.1 Introduction

In recent years, the Bullwhip Effect has garnered substantial interest in the supply chain management literature. Defined as the upstream magnification of demand variability, it translates into orders received upstream being more variable than the original customer demand, and generates inefficiencies in resource allocations and capacity, decreases in service level and customer satisfaction, as well as various other problems (Lee et al., 1997a). This results in higher costs for the supply chain and negatively impacts its profitability (Metters, 1997). In order to tackle the Bullwhip and its consequences, several solutions have been suggested, such as enhancing information visibility among partners, as well as better coordinating the supply chain operations (Lee et al., 1997b). Different types of collaboration initiatives have emerged, such as Collaborative Planning, Forecasting and Replenishment (CPFR) and Vendor-Managed Inventory (VMI), each aiming at ameliorating the performance of the supply chain (Holweg et al., 2005), with each having a different impact on the costs and profit of the upstream member (Kulp et al., 2004).

One of the suggested means to improve the information flow between supply chain members is the exchange of demand information in the form of Point-Of-Sales (POS) data, between the retailer and his upstream partners (Lee et al., 1997b). As customer demand is only observed by retailers, sharing demand information benefits upstream members by mitigating the detrimental impact of demand distortion due to the Bullwhip Effect. However, several challenges can arise in implementing this information scheme, such as the unwilligness of firms to cooperate, the lack of integration between data formats, implementation costs or confidentiality issues (Ali et al., 2017). In some cases, forecasts are exchanged instead of POS data, but these may be highly inaccurate and may not be shared truthfully by the supply chain members (Mishra et al., 2007, 2009). Conventionally, retailers do not gain any direct benefits from providing their POS data, but instead are offered incentives by their upstream partners to elicit the exchange of data (Lee and Whang, 1999; Cachon and Lariviere, 2001). Since information sharing serves as a mechanism for reducing the Bullwhip Effect and improving costs for the upstream member, understanding its benefits is of importance before embarking on any information exchange agreement.

There exists no general consensus on the value of information sharing in the literature. Indeed, conflicting results have emerged: some have advocated its use due to the benefits associated with it (e.g., Ali and Boylan, 2012), while others have argued against it, contending that upstream members are able to reconstruct the original demand signal from their incoming orders (for e.g., Graves, 1999). Various authors have studied this problem differently, as the mainstream approach has relied on using POS data exclusively to model information sharing (for e.g., Lee et al., 2000; Ali and Boylan, 2011), while a limited number of studies have employed both POS and retailer orders data (for e.g., Trapero et al., 2012; Williams et al., 2014). The univariate strategy, present in theoretical papers, effectively excludes the ordering policy and related information altogether from the forecasting model. This may discard pertinent features of the orders series that are absent from the demand data, such as ordering adjustments by managers. On the other hand, while the multivariate studies, which are empirically driven, have maintained that information sharing is beneficial for improving upstream forecasting accuracy, none of them has examined the inventory impact of deploying such a strategy, or made a direct comparison with the much simpler univariate POS approach.

Given the disparity in findings, as well as the difficulties associated with its implementation, information sharing remains a subject of debate among practitioners and academics. This paper aims at filling this gap in the literature by comparing various methods for modelling demand information sharing when forecasting uncertainty is present at all stages, and benchmarking their performance with the case of no information sharing (*NIS*). The following research questions are investigated in this paper: (i) What are the conditions under which information sharing is beneficial for the upstream member?; (ii) Which information sharing strategy brings most gains in terms of accuracy and inventory costs?; and (iii) What is the effect of ordering managerial interventions on information sharing as a whole and to each strategy?

Specifically, for a range of demand processes, we investigate two alternative multivariate approaches to the standard model for POS information sharing (which we refer to as univariate information sharing or *UIS*) in the upstream's forecasting model, which feature both historical orders and demand data. The first, the Multivariate Information Sharing or *MIS*, employs lags of the orders and demand series to predict retailer orders. The lags are restricted so that at the time of forecasting, all information is available (Sagaert et al., 2018; Ord et al., 2017). This means that for forecasting, for example, 3 steps ahead, the explanatory variables must have lag 3 or greater. This breaks any short-term connection between variables (Sagaert et al., 2018). To overcome this, we consider a second option, the Forecasting-based Information Sharing, *FIS*, which expands on the previous method by using forecasts of downstream demand to predict future orders and include them in the demand model, so as to access all lags of POS for all forecast horizons.

We conduct an inventory simulation to examine the impact of each strategy on forecasting accuracy, as well as their inventory performance. In addition, we assess the forecasting accuracy aggregated over lead-time, in stark contrast to previous research which relied on point forecast error metrics, as this is closer to the forecasts needed in practice. Furthermore, we introduce forecast uncertainty at all tiers in the supply chain, as well as managerial induced over-ordering at the downstream level to mimic practice, and contrast the value of information sharing with the case of no order deviations, which has been the standard assumption in most papers.

Overall, our results indicate that the traditional *UIS* model outperforms both *MIS* and *FIS* in terms of forecasting accuracy as well as total inventory costs, although the multivariate strategies offer improvements over the case of no information sharing. These help to reconcile findings in the literature, as they validate earlier results, while ranking the efficacy of each approach. The *FIS* method offers better results than *MIS*, however this does not hold for long manufacturer lead times. This can be understood as longer lead times entail a higher degree of forecast uncertainty, and since the *FIS* utilises short-term forecasts of customer demand, this uncertainty is further exacerbated, which is reflected in its subpar performance at higher lead times. Furthermore, when compared with the standard case of *NIS*, sharing information provides superior results on all the metrics in most of the cases, reinforcing the case for supply chain partners to communicate their demand data to improve their total performance.

The remainder of the paper is organised as follows. Section 4.2 first reviews the existing findings regarding demand information sharing, then covers the different strategies for modelling information sharing at the upstream echelon, before discussing the factors and circumstances under which information sharing would be favourable. Section 4.3 details the different information sharing methodologies studied in this paper, as well as the assumptions underpinning each strategy and their corresponding methodology. Section 4.4 describes the experimental setup for the simulation, while Section 4.5 analyses the results. Finally, Section 4.6 discusses the managerial implications of these findings.

4.2 Literature Review

In this section, we review the existing literature on information sharing. Section 4.2.1 first introduces the findings in favour of it (Section 4.2.1.1), then presents the results critical of it (Section 4.2.1.2). As the majority of these papers adopt a univariate approach to information sharing by relying exclusively on POS data, Section 4.2.2 proceeds to cover alternative multivariate information sharing approaches, which utilise both upstream and downstream demand information. Finally, Section 4.2.3 discusses the circumstances under which information sharing is expected to offer the most value in real-life settings, thus pointing out the difference between the assumptions made in theoretical papers and practical situations where information sharing would be necessary.

4.2.1 General Findings on Information Sharing

4.2.1.1 Arguments in Favour of Information Sharing

With the many deleterious consequences associated with the Bullwhip Effect, information exchange between supply chain members has been identified as a mechanism to mitigate them (Lee et al., 1997b). Various types of information can be shared such as: demand (Lee et al., 2000), projected orders (Zhao and Xie, 2002; Zhao et al., 2002; Chen and Lee, 2009), inventory (Cui and Shin, 2017; Srivathsan and Kamath, 2018), advanced demand information (Thonemann, 2002; Papier, 2016) or customer information (Yee, 2005; Li et al., 2019). This paper is concerned with information sharing in the form of retailers providing their upstream partners with the original customer demand.

As demand forecasting has been recognised as a source for the Bullwhip, upstream members can benefit by obtaining Point-of-Sales (POS) data as they have access to the original demand, unaltered by factors such as the effect of the retailer's inventory policy, forecasting or their lead time, which themselves contribute further to the Bullwhip Effect. The upstream members can improve their forecasting accuracy by using POS data, as they are able to use it to predict customer demand more accurately, instead of using solely the downstream orders they receive (Trapero et al., 2012). Consequently, lower safety stock levels could be held, given the reduction in forecasting errors. As a result, inventory holdings and their costs decrease (Lee et al., 2000; Yao and Dresner, 2008), as well as labour hours are more efficiently assigned (Sanders and Graman, 2016), which can increase the profitability of the upstream members and lead to an overall better performance of the supply chain (Yu et al., 2002).

Lee et al. (2000) studied information sharing in a two-stage supply chain with an AR(1) demand process, determining it to be of great benefit to upstream members as it leads to lower order variance, inventory reductions and cost savings. They also established that higher levels of positive autocorrelation in demand and longer lead-times make information sharing more valuable, as these factors increase the amplification of demand variability. Zhao and Xie (2002) found the demand process, as well as the distribution of the forecasting errors, to impact the value of savings from information sharing. Ali et al. (2012) showed that the magnitude of inventory reductions due to information sharing depend on the improvements in forecasting accuracy as a result of information sharing, while Ali et al. (2017) found that the inventory gains depend on the inventory cost structure. Sabitha et al. (2016) confirmed information sharing to be beneficial especially when lead times in the supply chain are decreasing. The inventory policy has also been determined to influence the value of information sharing (Cannella et al., 2011; Cannella, 2014; Costantino et al., 2015). Information sharing has been investigated in supply chains with more complex structures, with the results being favourable as it provides gains for the upstream members (Li, 2002; Cheng and Wu, 2005; Shang et al., 2015; Huang et al., 2017). However, despite these benefits, the Bullwhip Effect is alleviated but not completely eliminated, since it only tackles one contributing element of the phenomenon and does not remedy the other ones (Ouyang, 2007).

The advantages of demand sharing have also been verified by studies conducted empirically. Kelepouris et al. (2008) found that information sharing leads to reductions in order variability and higher fill rates for a warehouse serving one retailer. Hosoda et al. (2008) determined sharing of POS data to decrease the supplier's forecasting errors by up to 20%. Williams and Waller (2010) compared forecasts made by using POS data to those based on historical orders for a distribution centre, and concluded that the former provided a superior accuracy. Trapero et al. (2012) examined the impact of several forecasting methods for a manufacturer, including some with shared demand information, and established that information sharing returned the best forecasting accuracy, lowering the Mean Absolute Percentage Error (MAPE) by 6-8 % when compared to the case of no information sharing. Ali et al. (2017) observed reductions in both Mean Squared Error (MSE) and average inventory costs when the manufacturer resorted to information sharing. These empirical studies verify the claims of demand information sharing being beneficial to the upstream members from a practical perspective.

4.2.1.2 Arguments against Information Sharing

While POS exchange has been argued to offer benefits to the upstream members in the supply chain, some authors have determined its value to be limited under certain conditions (Raghunathan, 2003; Cavusoglu et al., 2012; Ganesh et al., 2014; Hartzel and Wood, 2017), while others have advocated for other measures such as reducing lead-times, as the latter would be be more effective (see for e.g., Cachon and Fisher, 2000; Steckel et al., 2004; Agrawal et al., 2009). Babai et al. (2016) proved that the inventory savings from information sharing for an AR(1) demand decrease at higher levels of demand autocorrelation. Teunter et al. (2018) highlighted that most papers suffered from a methodological flaw in the estimation of the variance of forecast errors, which lead to an overstatement of the benefits of information sharing.

Other bodies of research have adopted a more critical stance towards information sharing, suggesting that the latter brings little to no value for the upstream member (see for e.g., Graves, 1999; Raghunathan, 2001; Zhang, 2004a; Gilbert, 2005; Hosoda and Disney, 2006). The rationale behind this line of thought is that the demand information needs not to be shared, as it is already contained within the historical orders information. These theoretical papers share the common assumption of an ARIMA demand with known model and parameters, as well as a periodic Order-Up-To inventory policy at every stage. Therefore, a limited number of factors contribute to the upstream transmission of demand variability, which would render information sharing less advantageous, as fewer information distortions occur along the supply chain.

Graves (1999) studied a supply chain facing an ARIMA(0,1,1) demand and established that knowledge of the upstream demand parameters was only required in order to infer the customer demand. Raghunathan (2001) showed that for an AR(1) demand process the value of information sharing decreases as the number of historical orders increases, since the manufacturer can reconstruct the downstream demand from the orders more accurately. Gaur et al. (2005b) extended these results to ARMA(p,q) demand, and attribute this ability to infer demand from orders to the invertibility properties of the retailer's demand and order processes. They demonstrate that under certain assumptions, if the demand process is invertible with respect to its own shock, orders are inferrable from demand. Giloni et al. (2014) took their analysis further and argued that despite demand being invertible, the resulting orders may not be, and hence the manufacturer should inspect whether orders are invertible before deciding on information sharing. Kovtun et al. (2019) established that for a retailer observing two demand streams, each following an ARMA (p,q) representation, if the retailer's orders are invertible with respect to its own past shocks, sharing orders information (i.e. disaggregating the total order into the two components from which it is composed) offers the same results as sharing demand information; however demand information sharing is more valuable to the upstream member than order sharing.

The ability to infer demand from retailer's orders, which is contingent on the assumption of a known demand process, renders information sharing useless. Ali and Boylan (2011) addressed this issue, which they refer to as Dowstream Demand Inference (DDI), and explain that the assumptions made in those papers seldom hold in practice, and by removing these restrictions, they show that information sharing is beneficial to upstream members for an ARIMA(p,d,q) demand model with Minimum Mean-Squared Errors (MMSE) forecasts and an Order-Up-To inventory policy. They introduce two feasibility principles regarding DDI which depend on whether a unique translation of the downstream demand process to upstream orders exists: (i) if the upstream member is able to determine the demand process for the downstream member, they may not always be able to infer the demand values, and (ii) if the upstream member is unable to determine the downstream demand process, they may be able to reconstruct the demand values nonetheless. As a result, information sharing schemes where the model form and process are shared upstream, but the actual values are withheld, are no substitutes for full POS sharing, as the demand cannot be reconstructed by the upstream member. By relaxing this assumption, they prove analytically that information sharing results in higher levels of inventory reductions.

Ali and Boylan (2012) expanded on the previous research by studying the effect of non-optimal forecasting methods (in the form of Simple Moving Average and Simple Exponential Smoothing) for ARIMA demand. They determined that downstream demand may be inferred when the simple moving average forecasting method is implemented at the retailer level, which results in information sharing not being necessary. However, for the Simple Exponential Smoothing, demand cannot be inferred, and hence information should be shared to achieve cost reductions. Ali et al. (2017) showed that when DDI is applicable following the two feasibility principles, it still offers reductions in forecast errors and inventory levels, when compared to the case of no information sharing; nonetheless, these are higher when full information sharing takes place. Tliche et al. (2019) examined DDI for a causal invertible ARMA(p,q) process, determining information sharing to result in improvements in forecast accuracy reductions and decreases in inventory costs. These studies highlight the problems with inferring demand with orders, and strengthen the case for information sharing.

The majority of the aforementioned cases have investigated the advantages and disadvantages of information sharing, but they do not address the issue of how to incorporate the additional information in the forecasting process. Indeed, most of these papers have relied on the methodology featured in Lee et al. (2000), where the upstream members forecasts customer demand rather his incoming orders. This discards any information that might be contained in the order history but not in the demand data, which may prove to be relevant in the replenishment decisions of the upstream member. This raises the question of whether adopting a univariate strategy, by using solely the downstream demand information, is the most efficient way to represent information sharing, or whether featuring both sources of information from the orders and demand brings any additional benefits to the upstream member of the supply chain. Section 4.2.2 covers the different approaches to modelling information sharing at the upstream level.

4.2.2 Overview of Multivariate Methods in the Literature

Hanssens (1998) established that retailer orders and sales were cointegrated (i.e. they each exhibit a stochastic trend and share a co-movement in this trend) and displayed a long-term equilibrium. They used an Error Correction Model (ECM) to predict orders using sales and orders data. They found that it outperformed a univariate model fit on the historical data for orders in terms of forecast accuracy. Trapero et al. (2012) included historical lags of retailer sales and orders, modelling information sharing in two ways: a linear autoregressive model with exogenous sales inputs (ARX), and a nonlinear Artificial Neural Network and compared these two with various univariate time series model relying on orders data only. Models with POS data resulted in more accurate forecasts. Williams et al. (2014) found that customer sales and retailer orders exhibit an equilibrium in the long run due to the retailer's inventory policy, as the discrepancies between customer demand and orders placed to the distribution center are regulated in the short-run by the ordering decisions of the retailer. As a result, they construct a Vector Error Correction Model (VECM), and compare its forecasting performance with using solely POS data or retailer orders. The VECM model recorded a lower forecast error. In their empirical study, Cui et al. (2015) modelled information sharing by incorporating inventory information alongside lags of demand to capture the retailer's stock replenishment policy. They contrasted the forecasting performance in predicting retailer orders with three models: (i) a linear regression of orders on lags of customer demand, (ii) a linear regression of orders on lags of orders and lags of demands, analogously to the method featured in Trapero et al. (2012), and (iii) a Vector Autoregressive Integrated Moving Average (VARIMA), which treats orders and sales as endogenous variables. All three methods surpassed the benchmark of no information sharing in terms of forecast error, with their proposed method outperforming its counterparts.

All of the above-cited studies, which are based on real datasets, pinpoint the benefits of including both historical orders and sales over orders solely in the forecasting model, and overlook the inventory performance and its associated costs. Furthermore, all the aforementioned approaches are significantly more complex than using directly POS data. There are no widely accepted approaches to build either VECM or VARIMA models, raising the question of how much this additional complexity offers in terms of cost, especially since in practice the volume of forecasts would imply the need for a reliable automatic model specification methodology (Ord et al., 2017). As these results conflict with some of the theoretical findings opposing the exchange of information in a supply chain from Section 4.2.1.2, the circumstances under which information sharing is anticipated to offer value to upstream members should be better understood, as these might be omitted from the assumptions underlying the studies critical of information sharing. These are elaborated subsequently in Section 4.2.3.

4.2.3 Factors Affecting Information Sharing

One element that contributes to the upstream distortion of information is the uncertainty from forecasting demand. This is split into demand uncertainty, due to the inherent randomness in demand, and the uncertainty from estimating the forecasting model, due to mis-identifying the correct model or mis-specifying the parameters, omitting explanatory variables or over-specifying the model (Fildes and Kingsman, 2011; Saoud et al., 2019). In the studies where the demand process is known, the forecasting uncertainty is reduced to only demand uncertainty due to the stochastic errors, but in practice, determining the real demand process is difficult, which implies that the forecasting uncertainty will be large, which will generate higher costs for the supply chain. Zhao et al. (2002) determined that improvements in downstream forecasting accuracy enhance the value of information sharing, while Zhao and Xie (2002) studied the effect of forecast errors on the gains brought by information sharing and identified forecasting bias to be the main cost driver. Ali et al. (2012) quantified the impact of using sub-optimal forecasting methods for an ARIMA demand on the upstream performance when information is shared. Sanders and Graman (2016) established that forecasting bias at the downstream level was the most detrimental to the upstream member's total costs, and evaluated the relationship between forecast errors and information sharing. These studies all agree that the level of forecasting uncertainty influence the effectiveness of information sharing.

Given the prevalence of forecasting uncertainty, information sharing is expected to bring more advantages to the upstream members of the supply chain. Many characteristics of demand, such as seasonality, promotions and other special events make the extrapolation of future demand harder, exacerbating the effect of forecasting uncertainty. For instance, in the case study regarding the benefits of POS exchange undertaken by Kaipia et al. (2017), only products exhibiting these characteristics were selected by managers, since they were perceived to require information sharing to improve their performance. In the presence of seasonality, information sharing was determined to result in inventory and cost savings for the manufacturer (Zhao and Xie, 2002; Zhao et al., 2002; Cho and Lee, 2013). Retailer promotions, which were identified as one of the four original causes of the Bullwhip Effect (Lee et al., 1997b), also render the forecasting process more difficult, since surges in demand are hard to anticipate and incorporate in the forecasting model (Trapero et al., 2014). Sharing promotional plans with upstream members has been determined to increase the profitability of the manufacturer (Iyer and Ye, 2000; Tokar et al., 2011), as it allows the latter to better coordinate production and inventory decisions with the downstream promotional campaigns. In this case, other information can be shared between the supply chain members, such as the timing and type of the promotion, in addition to POS data, but this has not been explored yet in the literature.

As a result of the erratic nature of these events, managers often introduce adjustments at the forecasting level to better hedge against forecasting uncertainty (Trapero et al., 2013). Nonetheless, the efficiency of these adjustments has been questioned, as not all managerial alterations result in enhanced forecasting accuracy (Fildes and Goodwin, 2007; Fildes et al., 2009). These detrimental adjustments will be reflected in their subsequent ordering decisions, which will also impact their upstream partner. But apart from modifying the baseline forecasts, managers tend to change their orders to incorporate additional information not accounted for by the replenishment system output (Van Donselaar et al., 2010; Tokar et al., 2011, 2012). These adjustments can lead to over-ordering in some cases (Tokar et al., 2014), which contributes to the Bullwhip Effect and its associated costs (Croson and Donohue, 2005, 2006).

Many behavioural studies have identified several reasons for managers to deviate from the optimal ordering decisions, including: under-estimating the order pipeline (Sterman, 1989; Niranjan et al., 2011), anchoring bias (Schweitzer and Cachon, 2000; Kremer et al., 2010), censorship bias (Feiler et al., 2013), overconfidence (Ren and Croson, 2013), the durability of inventory and its transit lags (Bloomfield and Kulp, 2013), bounded rationality (Su, 2008), over-estimating the value of substitute products (Bansal and Moritz, 2015), or even psychological and behavioural reactions to spikes in demand (Sterman and Dogan, 2015). Trapero et al. (2012) and Cui et al. (2015) maintain that the omission of these deviations in the supply chain studies is one of the reasons why some theoretical papers derive no benefits from information sharing. We conjecture that including these aspects should introduce more distortions to the orders placed, and thus make the forecasting process upstream more difficult. We further conjecture that because of these deviations, the POS data should be more useful as it allows the upstream members to observe the initial customer demand, unaltered by these distortions.

4.3 Information Sharing Modelling Strategies

As mentioned previously, this paper aims at comparing different strategies for sharing information. In this section, we present these approaches, explaining their underlying rationale as well as their modelling methodology. We restrict our attention to a supply chain consisting of a single retailer and manufacturer. The retailer observes the downstream demand D_t at time t, and places his orders O_t to the manufacturer, which serve as the latter's demand signal. The supply chain is decentralised, in that despite the possibility of information exchange, each member plans their respective ordering and inventory decisions independently. Thus, the difference between each of the methods discussed below boils down to which information source is employed when building the manufacturer's forecasting method, as well as how it is modelled.

Four different strategies are considered. The first is the standard case of No Information Sharing (*NIS*), where the manufacturer uses only historical retailer orders as inputs to forecast their incoming orders. The second is the Univariate Information Sharing approach (*UIS*), which is the traditional method for information sharing, where the manufacturer employs historical POS data exclusively to build the forecasting model. Both *NIS* and *UIS* are univariate methods as they depend only on one source of information (retailer orders and POS data respectively). The third is the Multivariate Information Sharing *MIS*, which utilises both O_t and D_t as inputs to the forecasting model. However, not all lags of D_t are used in *MIS*; instead only lags of demand exceeding the forecast horizon feature, in order to avoid the need to extrapolate D_t. In order to overcome this possible limitation, a fourth strategy is tested, which produces forecasts for these discarded lags of D_t. We denote this approach as Forecasting-based Multivariate Information Sharing Method (*FIS*). In order to better understand the difference between each method, as well as their implementation, next we detail their underlying methodology. All the above-mentioned methods use the ARIMA family of models as their building block (Box et al., 2015).

The first step consists of checking whether the retailer orders and customer demand series are stationary. Following the Box and Jenkins methodology (Box et al., 2015), as well as to avoid spurious regressions (Ord et al., 2017), the series are tested and transformed to attain stationarity. To detect the presence of a unit root, we use the (KPSS) test (Kwiatkowski et al., 1992). To identify seasonality, we test whether the distribution of at least one of the potential seasonal patterns is statistically significant from the rest, relying on the non-parametric Friedman test (Hollander et al., 2013). The series are then differenced accordingly.

After rendering the series stationary, the univariate forecasting methods (NIS and UIS) are fitted. The methods with univariate information are simpler to model, as they do not depend on external variables (Sagaert et al., 2018). Given that the previous steps guarantee stationarity for the processes, the NIS and UIS are thus modelled as ARMA(p,q). To determine the ARMA order, the search strategy by Hyndman and Khandakar (2008) is adopted, starting with an ARMA(2,2) model and searching the model neighborhood by modifying the order for p and q by ±1. For each model, the in-sample Akaike Information Criteria corrected for small sample sizes (AICc) is recorded (Burnham and Anderson, 2002), and if a better model is identified, then the search restarts from there. This process is repeated until no better model can be identified. In order to avoid over-specified models, the maximum order for p and q are restricted to 5 lags. At this stage, the upstream demand (O_t) is used to build the ARMA model for the *NIS*, *MIS* and *FIS* approaches, while for the *UIS*, the downstream demand D_t is employed. The modelling for the univariate strategies (NIS and UIS) ends here.

The forecasts are produced and then differencing is reversed to obtain the final predictions.

For the multivariate approaches (MIS and FIS), the next step consists of incorporating the downstream demand in the ARMA model. Once the latter is fit, the residuals are then extracted, $q_t = z_t - \hat{z}_t$, where $\hat{z}_t$ are the ARMA predictions and z_t are the stationary observations. We construct autoregressive lags up to $H+1$ for the POS data (where H is the manufacturer's maximum horizon) and a linear regression is fitted on q_t. The two-step specification, where first the univariate ARMA is built and then the POS information is used, is employed to ensure that each information group is used effectively (Ma et al., 2016), and that univariate information is preferred. The appropriate lags are selected using a stepwise regression, by evaluating the contribution of each input in terms of AICc. We restrict the lags so that the information is always conditioned correctly on the forecast origin, i.e. if we want to predict 3-steps ahead, then lags $t-1$ and $t-2$ are not permitted, since these would necessitate information from the future (Sagaert et al., 2018). One way to overcome this limitation is to produce forecasts on the explanatory variable (POS data), so as to obtain values for the future periods (Ord et al., 2017). This constitutes the difference between MIS and FIS: the former uses restrictions on the lags, while the latter uses forecasts of the stationary POS series and therefore allows all lags to be present in the regression equation. The final prediction is the sum of the ARMA forecast and the regression on the residuals, which is then undifferenced to achieve the final forecast.

4.4 Experimental Design

4.4.1 Data Simulation

Our simulation consists of a serially linked decentralised dyadic supply chain consisting of a manufacturer and a retailer selling a single product. At time t, the retailer observes customer demand D_t, receives the shipment for the orders placed h periods ago (where the forecasting h is defined as the sum of the lead time and the inventory review period), produces future forecasts over the horizon, and places the orders based on the safety stock requirements and current inventory position. Unmet demand is backlogged. Similarly, the manufacturer follows a similar sequence of events where they forecast the received orders from the retailer (which now serve as their demand), observe their inventory, and places orders to a supplier assumed to have sufficient inventory to meet all demand. No capacity constraints are imposed on the manufacturer.

Process	Data Generating Process
AR(1)	$(1-\phi B)D_t = \varepsilon_t$.
IMA(1,1)	$(1-B)D_t = (1+\theta B)\varepsilon_t$.
ARIMA(0,1,1)(0,1,1)$_{12}$	$(1-B)(1-B^{12})D_t = (1+\theta B)(1+\Theta B^{12})\varepsilon_t$

Table 4.1: Data Generating Processes for demand.

Three data generating processes (DGPs) are investigated in this paper, each belonging to the ARIMA family of models (Box et al., 2015). These are displayed in Table 4.1, where B denotes the backshift operator, defined as $B^n(D_t) = D_{t-n}$, and $\varepsilon_t \overset{i.i.d}{\sim} N(0,\sigma^2)$. The first is the well-known stationary AR(1) process, which has been employed in several papers (see for e.g.„ Lee et al., 2000; Raghunathan, 2001; Ali and Boylan, 2011); the second (Graves, 1999) contains a stochastic unit root, and corresponds to the Simple Exponential Smoothing model (Gardner Jr, 2006); while the final model encompasses both a stochastic trend and seasonal pattern, and is equivalent to the Airline Passengers model (Box et al., 2015). For each process, the model parameters are drawn to ensure stationarity and invertibility (Ord et al., 2017).

To guarantee positive demand observations, a constant level is added to all time series. The standard deviation of the innovation terms is: $\sigma \in \{1,5,10\}$, representing different levels of demand volatility. Under each setting 500 observations are generated for the demand series, from which the first 200 points constitute the training set over which the forecasting model and parameters are estimated, the next 200 points are used to allow the inventory policy to warm-up, eliminating any bias from its initialisation, and the last 100 points represent the test set over which the results are calculated. In total, 1000 replications for each simulation setting are produced, generating 9000 retailer demand series that result in 1.458 million cases for the manufacturer.

We do not assume knowledge of the DGP. The retailer adopts an automatically specified ARIMA following the methodology by Hyndman and Khandakar (2008). As a result, the retailer's forecast may be mis-specified, reflecting increasing uncertainty in the supply chain. The retailer's forecasts are transformed to orders using an adaptive Order-Up-To (OUT) inventory policy. This generates the demand that the manufacturer observes, which is modelled with the alternatives outlined in Section 4.3.

The horizons for the retailer and manufacturer, h and H respectively, are $\{1,3,5\}$. This permits us to study the lead-time effect on the value of information sharing. The cycle service level for the OUT policy for both retailer and manufacturer is set at $\alpha \in \{90\%, 95\%, 99\%\}$, to mirror values employed in practice. All point forecasts are produced according to a rolling origin forecasting scheme, with the model and parameters estimated only once over the training set. Similarly to previous research, we assume that all unfulfilled demand is backordered, and no set-up or fixed ordering costs exist. The Order-Up-To level S_j at time index j is determined as: $S_j = \sum_{j=1}^{h} \widehat{D}_j + k_\alpha \widehat{\sigma}_j$, where $\sum_{j=1}^{h} \widehat{D}_j$ is the cumulative demand forecasts aggregated over the horizon h, k_α is the inverse from the cumulative normal distribution for cycle service level α, and $\widehat{\sigma}_j$ is the cumulative conditional standard deviation of the horizon errors (Saoud et al., 2018). For both retailer and manufacturer, the inventory review period is 1. At each iteration, the $\widehat{\sigma}_j$ estimate is updated to include the available forecasting errors. The initial inventory level is initialised by setting it equal to its safety stock level over its horizon.

An argument in favour of more complex options than simply using *UIS* has been that *UIS* does not capture any information in the supply chain beyond the retailer's demand. To explore this, we introduce disruptions between the retailer and the manufacturer. Therefore, we consider two different settings: the first is the supply chain operating as is, and the second is the supply chain in the presence of managerial adjustments amounting to over-ordering, a phenomenon reported in previous research (see Section 4.2). This is achieved by introducing a deviation term δ_i to the final orders generated by the inventory policy of the retailer. This term is defined as: $\delta_i = s p_i \xi_i$, where s is the standard deviation of the (stationary) demand series, p_i an indicator variable that is equal to 1 when a random value drawn from $U(0,1) \leq 0.3$ and otherwise zero, and $\xi_i \sim \Gamma(2, 0.5)$ to ensure positive draws. The parameters for the Gamma distribution were experimented with to produce reasonable ordering deviations. An example is provided in Figure 4.1. We produce 1000 replications for each setting, which generates 9000 retailer demand series that result in 1.458 million cases for the manufacturer. The full set of control parameters for the inventory simulation are tabulated below in Table 4.2

Variable	Values	Options
Information Sharing Method	NIS, PIS, MIS, FIS	4
Downstream Demand Process	AR(1), IMA(1,1), ARIMA(0,1,1)(0,1,1)$_{12}$	3
Demand Noise Level	1, 5, 10	3
Retailer Horizon	1, 3, 5	3
Manufacturer Horizon	1, 3, 5	3
Retailer Service Level	90%, 95%, 99%	3
Manufacturer Service Level	90%, 95%, 99%	3
Ordering Adjustment Frequency	0, 0.3	2

Table 4.2: Experimental Design Control Parameters

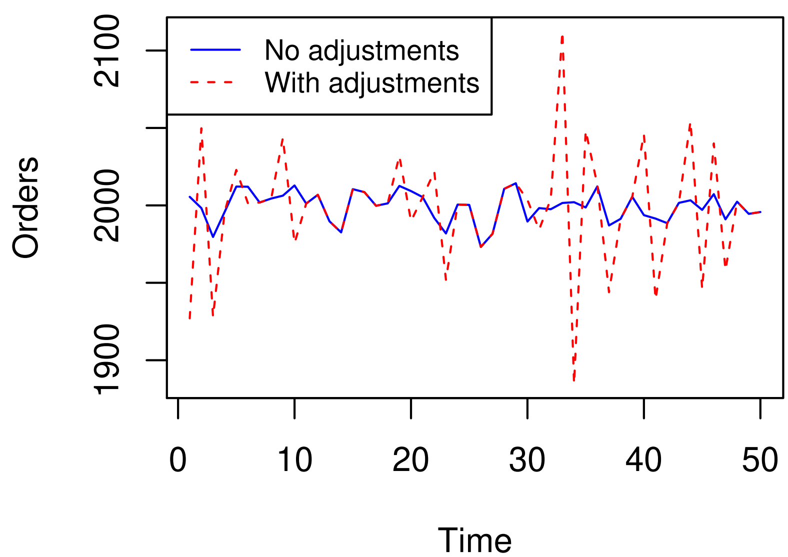

Figure 4.1: Example of retailer orders with and without order adjustments for an AR(1) demand process. Observe that the ordering pattern becomes more erratic due to excess stocking from the adjustments.

4.4.2 Evaluation Metrics

We assess the performance of the different information sharing methods by looking at their forecasting accuracy and inventory costs. In this paper, forecasting accuracy is measured using the Ratio of Forecasting Uncertainty (RFU) proposed by Saoud et al. (2019). This consists of the ratio of manufacturer to retailer's cumulative Root Mean Squared Error over their respective horizons, with the latter defined as the conditional standard deviation of the forecast errors aggregated over the horizon that is used in the determination of safety stock levels. More specifically, the cumulative RMSE can be written as:

$$\text{CumRMSE} = \sqrt{\frac{1}{(N-H+1)} \sum_{t=1}^{N-H+1} \left(\sum_{i=1}^{H} D_{t+i} - \sum_{i=1}^{H} \widehat{D}_{t+i|t} \right)^2} \tag{4.1}$$

with N and H denoting the number of observations and forecasting horizon, and $\widehat{D}_{t+i|t}$ the i-th step-ahead point forecast made for demand D_{t+i} at period t. The summations $\sum_{i=1}^{H} D_{t+i}$ and $\sum_{i=1}^{H} \widehat{D}_{t+i|t}$ calculate the cumulative demand and forecast, while the rest of the formula calculates the RMSE over the sample. The RFU is determined as:

$$\text{RFU} = \frac{\text{CumRMSE}^M}{\text{CumRMSE}^R} \tag{4.2}$$

where the subscripts M and R refer to manufacturer and retailer respectively. This metric incorporates the forecasting accuracy of both members of the supply chain, and evaluates how forecasting uncertainty is evolving as we move upstream in the supply chain. In addition, it overcomes many of the limitations of the standard measure for the Bullwhip Effect and has been found to display a better relationship with inventory costs (Saoud et al., 2019). For the *UIS* method, the manufacturer will calculate their CumRMSE with respect to the customer demand. Comparing that value to the CumRMSE of the retailer, these are expected to be very close, as they use the same data. However, differences can appear due to the set of data used for training by each member or the difference in forecasting models employed for example.

To gauge the stock control performance, we measure total inventory costs for the manufacturer, defined as the sum of the backorder costs c^- and holding costs c^+. Given the difficulty in estimating the inventory costs, we use the following approximation for cycle service level α:

$$\alpha \approx \frac{c^-}{c^- + c^+} \tag{4.3}$$

The value of c^+ is set at 1 monetary unit and $c^- \in \{9, 19, 99\}$.

For each of the metrics under study, the ratio of geometric means is computed with respect to the benchmark of no information sharing. The underlying idea behind the use of geometric means instead of arithmetic means is that it is more suitable when there might exist a scale issue between the observations, which is expected due to the different settings resulting in heterogeneous results, thus allowing a direct scale-independent comparison of the information sharing methods to *NIS* (Fleming and Wallace, 1986). More specifically,

$$\widetilde{X}_S^{(IS)} = \sqrt[n]{\frac{\prod_{j=1}^{n} X_{j,S}^{(IS)}}{\prod_{j=1}^{n} X_{j,S}^{(NIS)}}} \tag{4.4}$$

where $X \in$ {RFU, Total Cost} and IS can be any of the information sharing method (*UIS*, *MIS* and *FIS*), n is the number of replications under simulation settings S. A similar approach can be found in the works by Fildes (1992) and Davydenko and Fildes (2013), who provide an explanation for the rationale behind the use of geometric means of ratios in a forecasting accuracy context. This ratio is easy to interpret, as values above 1 indicate that the method fails to surpass the benchmark of *NIS*, and vice versa. The simulation was conducted using the *R* statistical language (R Core Team, 2019).

4.5 Results

Given the large number of dimensions in the experimental study, we first inspected which settings affected the result rankings. The service levels and the standard deviation of the DGP did not alter the rankings of the different information sharing methods. Therefore, we average across these dimensions and retain the DGP type, the horizons and the order deviation as pertinent variables. The simulation results are first presented for the case of no order adjustments, and then contrasted with those where the deviations are introduced.

4.5.1 No Adjustments

4.5.1.1 Forecasting Performance

The results for the RFU values with no ordering adjustments are displayed in Table 4.3. The table is organised as follows. Rows are grouped in 3 sets, one for each triplet of *UIS*, *MIS* and *FIS* per DGP. Columns are organised by retailer (h) and manufacturer (H) horizons. The best performance per set and horizon is highlighted in boldface. As can be observed from the table, the UIS method has a better forecast accuracy than no information for all combinations of horizons for the AR(1) process, as indicated by the values for its ratios being less than 1, thus agreeing with the previous results from the literature on the forecasting gains from using demand information for the manufacturer. Both the *MIS* and *FIS* approaches

Process	Information Sharing	Retailer Horizon 1			Retailer Horizon 3			Retailer Horizon 5		
		Manufacturer Horizon			Manufacturer Horizon			Manufacturer Horizon		
		1	3	5	1	3	5	1	3	5
AR(1)	UIS	0.92	**0.89**	**0.94**	**0.67**	**0.84**	**0.85**	**0.58**	**0.78**	**0.86**
	MIS	0.99	1.01	1.01	0.97	0.98	1.01	0.72	0.86	0.8
	FIS	**0.89**	0.99	0.99	0.73	0.85	0.98	0.68	0.83	0.88
IMA(1,1)	UIS	0.97	**0.85**	**0.87**	**0.57**	**0.69**	**0.68**	**0.39**	**0.50**	**0.59**
	MIS	0.99	1.01	1.01	0.96	0.97	1.03	0.56	0.91	1.12
	FIS	**0.93**	0.99	0.99	0.65	0.89	1.04	0.54	0.92	1.21
ARIMA (0,1,1)(0,1,1)	UIS	**0.72**	**0.81**	**0.86**	**0.35**	**0.48**	**0.57**	**0.24**	**0.33**	**0.13**
	MIS	1.01	0.99	0.99	0.99	0.99	0.99	0.76	0.89	0.64
	FIS	0.90	0.98	1.01	0.80	0.95	1.11	0.76	0.89	1.03

Table 4.3: RFU ratios for the experimental settings under no retailer ordering deviations

also present a favourable forecasting performance, as the ratios are mostly below 1 across all horizons. This implies that the inclusion of the downstream demand in the forecasting model augments its performance, as relevant information is contained in the POS data, which allows a better prediction of the incoming orders at the manufacturer level, and reinforces the case for information sharing between supply chain members. Nonetheless, the univariate method dominates the multivariate one in terms of forecasting accuracy, despite the *FIS* method offering slightly superior results when both horizons are set at 1. Contrasting both multivariate approaches (*MIS* and *FIS* respectively), we observe that the latter outperforms the former, implying that the introduction of future forecasts for customer demand in the upstream forecasting model ameliorates its performance over their omission. This can be understood as *MIS* relies on long lags to extract any relevant information, while *FIS* has access to information from shorter lags, albeit forecasts and not realised values. With regards to the effect of the forecast horizon, it can be seen that higher values for the retailer's horizon enhance the value of information sharing for the manufacturer, as it entails a higher level of forecasting uncertainty downstream, which would be exacerbated upstream, and as a result rendering information sharing more beneficial for the manufacturer.

The findings for the AR(1) process are mirrored by the non-stationary processes, although the magnitude of the results differ. Indeed, the univariate and multivariate methods offer promising performance in the case of the IMA(1,1), outperforming the case of no information sharing, with the *UIS* surpassing its counterparts. However, the values for the ratio of the *UIS* are lower when compared to the AR(1), implying that information sharing offers more advantages for this process. This does not however apply to the multivariate information sharing strategies, as higher horizons for the manufacturer are associated with an increase in forecasting uncertainty (as manifested by higher RFU values) when compared to those for the AR(1) demand process. The *FIS* also manages to surpass the *MIS* in most cases (as well as the *UIS* for $h = H = 1$), reinforcing the argument made for the AR(1) results. These results can be anticipated as non-stationary demand processes are harder to forecast than stationary ones, and hence information sharing helps alleviate the forecasting uncertainty better than for the stationary process.

For the seasonal demand process, the performance of the different information sharing strategies for the ARIMA(0,1,1)(0,1,1)$_{12}$, which has not been studied in the literature, are similar to those of the IMA(1,1). Indeed, all three approaches offer gains in forecasting accuracy, as evidenced by the ratios being less than 1. When comparing those values to the AR(1) results, it is clear that information sharing for the seasonal process brings greater reductions in forecasting uncertainty, which can be observed especially in the case of $h = 5$, where the ratio takes values between 0.13 and 0.33, reflecting a much superior forecasting accuracy than *NIS*. Both multivariate methods again display a preferable forecasting performance than no information sharing, and the decrease in uncertainty resulting from the exchange of demand information for the seasonal process exceeds that for the other non-seasonal processes. As seasonality renders the forecasting process even harder due to the additional uncertainty from estimating this component, information sharing appears to provide higher gains in forecasting accuracy for the manufacturer, as the propagation of forecast uncertainty is even higher.

The results so far confirm the findings in the literature that *UIS* and *MIS* like approaches are beneficial; in addition, the *FIS* strategy, which generates short-term forecasts for customer demand in order to extrapolate incoming orders, provides preferable results over *MIS*; however we find that *UIS* is generally superior, as it uses data with much lower variability to produce the forecasts, while the conditional expectations for both the retailer and the manufacturer remain the same.

4.5.1.2 Inventory Performance

With all three information sharing strategies having displayed good forecasting accuracy, we now turn our attention to the inventory results to inspect whether the gains carry across to cost reductions for the manufacturer. The results are tabulated in Table 4.4, which follows the same structure as Table 4.3. The improvements in forecasting accuracy from the univariate and multivariate methods for information sharing are reflected in their inventory performance. Indeed, for the AR(1) process, all three methods display lower costs then no information sharing, with the univariate strategy again returning the best inventory performance. Furthermore, these decreases in costs are especially pronounced when the manufacturer's lead time is zero ($H = 1$). As the retailer's horizon increases, information sharing becomes more valuable in diminishing inventory costs, especially for the *MIS* method where the values of the ratio decreases. Hence, for higher retailer horizons, information sharing proves to be more valuable for all three strategies. As for the *FIS*, it again fails to beat the benchmark of *NIS* in any situation, and this is attributed again to the compounding of the forecasting uncertainty due to the inclusion of forecasts for the POS data.

Process	Information Sharing	Retailer Horizon 1			Retailer Horizon 3			Retailer Horizon 5		
		Manufacturer Horizon			Manufacturer Horizon			Manufacturer Horizon		
		1	3	5	1	3	5	1	3	5
AR(1)	UIS	0.85	**0.79**	**0.89**	**0.46**	**0.71**	**0.73**	**0.35**	**0.6**	**0.74**
	MIS	0.99	0.99	1.01	0.95	0.97	1.02	0.53	0.74	0.8
	FIS	**0.81**	0.97	0.99	0.55	0.72	0.96	0.47	0.70	0.78
IMA(1,1)	UIS	0.93	**0.72**	**0.76**	**0.33**	**0.47**	**0.46**	**0.16**	**0.25**	**0.33**
	MIS	0.96	0.96	0.96	0.88	0.88	1.02	0.31	0.73	1.12
	FIS	**0.85**	0.93	0.95	0.42	0.72	0.99	0.29	0.74	1.25
ARIMA (0,1,1)(0,1,1)	UIS	**0.61**	**0.73**	**0.81**	**0.23**	**0.23**	**0.3**	**0.11**	**0.09**	**0.13**
	MIS	0.97	0.92	0.92	0.86	0.77	0.76	0.47	0.54	0.64
	FIS	0.87	0.89	0.96	0.58	0.68	0.92	0.47	0.55	0.66

Table 4.4: Total cost ratios for the experimental settings under no retailer ordering deviations

For the non-stationary processes, we again observe that information sharing offers more benefits for the manufacturer, again falling in line with the forecasting accuracy findings and those for the AR(1) inventory costs. Comparing the results of the seasonal process with the IMA(1,1) show that in the presence of seasonality, all information sharing methods methods again are more advantageous, as the difference between either approach and the no information sharing setting further widens. However, for high manufacturer horizons, the multivariate methods return higher costs than no information sharing, unlike their univariate counterpart. Comparing the performance of both *FIS* and *MIS*, we observe that the latter offers lower total cost ratios than then the former in many cases, thus making it a slightly preferable option for the manufacturer. Again, the univariate method displays the lowest cost ratios, and all of these findings closely match those from Table 4.3, as the accuracy results and rankings are mirrored in the inventory performance of each method.

4.5.2 Ordering Adjustments

The above results pertain to the case where no managerial intervention occurs before placing the final orders. We now study the value of information sharing in the presence of these adjustments, where arguably *MIS* should have an advantage. The accuracy results are tabulated in Table 4.5, which is organised as Table 4.3.

4.5.2.1 Forecasting Performance

Process	Information Sharing	Retailer Horizon 1			Retailer Horizon 3			Retailer Horizon 5		
		Manufacturer Horizon			Manufacturer Horizon			Manufacturer Horizon		
		1	3	5	1	3	5	1	3	5
AR(1)	UIS	**0.23**	**0.48**	**0.62**	**0.23**	**0.46**	**0.59**	**0.22**	**0.45**	**0.74**
	MIS	1.01	1.01	1.01	1.01	1.01	1.02	0.93	0.94	0.8
	FIS	0.92	0.99	1.06	0.91	0.93	1.05	0.89	0.93	0.98
IMA(1,1)	UIS	**0.22**	**0.43**	**0.55**	**0.21**	**0.39**	**0.49**	**0.19**	**0.34**	**0.33**
	MIS	1.01	1.03	1.05	0.99	1.02	1.04	0.86	1.12	1.12
	FIS	0.93	1.21	1.59	0.9	1.17	1.66	0.85	1.15	1.59
ARIMA (0,1,1)(0,1,1)	UIS	**0.2**	**0.35**	**0.44**	**0.16**	**0.3**	**0.39**	**0.14**	**0.25**	**0.13**
	MIS	0.99	1.02	1.03	0.99	1.02	1.03	0.91	0.98	0.64
	FIS	0.97	1.07	1.19	0.94	1.04	1.2	0.91	0.99	1.13

Table 4.5: RFU ratios for the experimental settings with retailer ordering deviations

The relative RFU for the different methods differ from the previous case of no order adjustments for all processes. Indeed, the univariate method produces much lower values when compared to the previous case. For *MIS*, we notice that the results are minimally affected by the introduction of these adjustments, as the values for the ratio are very close to those of *NIS* (with the exception of the case of high retailer horizon, where the *MIS* offers a better forecasting performance than no information sharing). Hence, including the additional POS information in the forecasting model does not help the performance of this method when ordering adjustments occur. As for the *FIS*, while it still displays the lowest ranking across all four methods, we observe a slight improvement for the AR(1) results, but overall it does not bring any value over no information sharing. Note that the rankings of the methods does not change when compared to the case of no managerial adjustments, despite the relative ratios differing in size. We verify this claim by calculating at the correlation between the results for no order adjustments against order adjustments, which is found to be 0.75. It is interesting to understand why *UIS* remains the top performer while both *MIS* and *FIS* are only marginally helpful. Although the adjustments are strictly positive, the inventory policy will in the long term retain the mean level of demand at the right level (see for example, Figure 4.1). Therefore, the equality of conditional expectations of D_t and O_t still holds and hence the less variable D_t that is exclusively used by *UIS* brings the maximal gains.

4.5.2.2 Inventory Performance

Process	Information Sharing	Retailer Horizon 1			Retailer Horizon 3			Retailer Horizon 5		
		Manufacturer Horizon			Manufacturer Horizon			Manufacturer Horizon		
		1	3	5	1	3	5	1	3	5
AR(1)	UIS	**0.23**	**0.48**	**0.62**	**0.23**	**0.46**	**0.59**	**0.22**	**0.45**	**0.74**
	MIS	0.99	1.01	1.01	1.01	1.02	1.04	0.87	0.86	0.8
	FIS	0.84	0.97	1.07	0.83	0.85	1.06	0.8	0.85	0.91
IMA(1,1)	UIS	**0.22**	**0.43**	**0.55**	**0.21**	**0.39**	**0.49**	**0.19**	**0.34**	**0.33**
	MIS	0.92	0.94	1.03	0.9	0.91	0.97	0.66	1.01	1.12
	FIS	0.79	1.22	1.96	0.73	1.14	2.11	0.64	1.06	1.89
ARIMA (0,1,1)(0,1,1)	UIS	**0.2**	**0.35**	**0.44**	**0.16**	**0.3**	**0.39**	**0.14**	**0.25**	**0.13**
	MIS	0.77	0.63	0.61	0.74	0.59	0.56	0.59	0.5	0.64
	FIS	0.74	0.68	0.77	0.66	0.59	0.71	0.59	0.51	0.57

Table 4.6: Total cost ratios for the experimental settings with retailer ordering deviations

The insights from the case of no order deviations are echoed at the inventory level as well in Table 4.6. All information sharing methods display lower inventory costs than *NIS*, although these reductions are greater in the presence of the managerial adjustments. For instance, the relative cost ratios for the *UIS* are substantially lower, which bolster the case for information exchange between the supply chain members. The *MIS* and *FIS* methods also offer decreases in inventory costs, which exceed those from the case of no ordering adjustments. Furthermore, as the demand process becomes more complex or when the horizon increases, this strategy proves to be more beneficial for the manufacturer, again agreeing with the insights derived previously. Given that this setting reflects the situation faced by upstream decision makers better than the previous case of no adjustments, we thus find that information sharing is even more advantageous from an inventory perspective. The *FIS* method overall produces substandard results for long manufacturer lead times, despite displaying cost improvements over no information sharing in some instances for the stationary AR(1). The inventory cost rankings are again unaffected by the introduction of these adjustments (correlation is 0.672). This might appear counter-intuitive at first, but as we argued for in the forecasting results, the inventory policy ensures that the conditional expectations remain the same and thus the results remain broadly the same.

4.6 Conclusion

Information sharing has attracted considerable interest from researchers and practitioners alike, as it allows for upstream members in the supply chain to gain visibility into the downstream demand, who in turn can potentially attenuate the negative consequences of the Bullwhip Effect and decrease their related costs. While the bulk of the literature has focused on understanding whether such a scheme is indeed beneficial or not, this paper is concerned with the forecasting modelling strategies that may be used to represent the demand data. Specifically, we conduct a comparison between the standard univariate method for information sharing (*UIS*), which has featured in the majority of studies, and two other multivariate approaches which rely on both orders and sales data (*MIS* and *FIS*), with each built on a different set of modelling assumptions. Furthermore, these are assessed with respect to a benchmark of no information sharing, *NIS*, to determine crucially whether information sharing is beneficial or not from a forecasting and inventory perspective.

The performance of the aforementioned models is evaluated through a supply chain simulation in terms of forecast accuracy, and inventory holding costs. Our results indicate that while both the univariate and multivariate approach display a higher forecasting accuracy and lower inventory costs than in the case of no information sharing, the former outperforms its counterparts across all settings, thus making it the most promising strategy out of the four evaluated.

The simulation indicates that the value of information sharing increases when: (i) the demand process becomes more complex (from stationary to nonseasonal non-stationary, to seasonal non-stationary), (ii) when the retailer horizon increases and (iii) when deviations from the recommended ordering level are introduced, matching better the realistic case. Our results agree with the literature in favour of sharing information for the AR(1) process (for e.g., Lee et al., 2000; Ali and Boylan, 2012) and the IMA(1,1) process (Babai et al., 2013), and indicate that for the ARIMA(0,1,1)(0,1,1)$_{12}$ process, the exchange of information is of greater value than for the other cases. Furthermore, we demonstrate that the gains are reflected in the inventory costs in all scenarios.

We argue that *UIS* remains the best model, even after the introduction of managerial ordering adjustments, due to the inventory policy accounting for any induced overstocking. The conditional expectations of retailer and manufacturer demand will remain the same, and therefore the reduced variability of *UIS* provides the most benefits. Both *MIS* and *FIS* are found to provide advantages over *NIS*, as reported in the literature on multivariate information sharing, but not to the same extent as *UIS*. This could possibly be inversed in a case where a potentially more simplistic inventory policy would not account for the effect of managerial ordering adjustments in the long term. This might reflect more accurately some of the current practice, and is a limitation of this work. For instance, in the work by Cui et al. (2015), the retailer adheres to a Constant Day of Inventory policy, as opposed to the OUT policy, and information sharing is deemed to be more valuable in the presence of the aforementioned ordering deviations.

Finally, we draw the reader's attention to the formulation of *MIS*. The POS data was conditionally restricted, i.e. the minimum lag could not exceed the forecasting horizon. This may weaken the connection between POS and manufacturer demand data. We attempted to overcome this with the evaluation of the *FIS* method, which displayed an improvement over both *NIS* and *MIS* for the majority of lead times. We attribute this performance to the ability of the *FIS* method to utilise information found in short lags in its forecasting model. However, at longer lead times, we observe a deterioration in its performance due to the increase in forecast uncertainty over long lead times and the accumulation of forecast errors, especially for the non-stationary series. Different forecasting model selection schemes, or other sources of information that extend beyond POS data, or even alternative DGPs, may be more favourable to *MIS*, and should be considered. Nonetheless, this study provides a thorough evaluation for a number of scenarios using widely recognised approaches to automatically specify models for *MIS*.

Acknowledgments

All calculations were performed on the High End Computing Facility at Lancaster University, United Kingdom.

Chapter 5

Conclusion

In recent years, the Bullwhip Effect has been identified as a "hot topic" for research in Supply Chain Management, due to the many inefficiencies and unnecessary costs it generates for supply chains (Wang and Disney, 2016). Numerous studies have been dedicated to better understanding the causes of this phenomenon, its impact and prevalence in various sectors, together with possible solutions to mitigate or eliminate it. One factor that has been highlighted as causal is the quality of demand forecasts. Forecast uncertainty, as measured by forecast errors, amplifies upstream along the supply chain (Lee et al., 1997a). Studying the uncertainty resulting from the forecasting process is crucial to better understanding the impact of the Bullwhip Effect on the Supply Chain.

Throughout this thesis, three key research questions were investigated: (i) How can we adequately capture forecast uncertainty in a supply chain? (ii) How can we measure the upstream propagation of forecast uncertainty resulting from the Bullwhip Effect? (iii) What downstream demand information sharing strategy is the most effective in reducing forecast uncertainty and upstream inventory costs resulting from the Bullwhip Effect?

In Chapter 2, we explored how to measure forecast uncertainty for inventory purposes at the retailer level, where the Bullwhip Effect is triggered. Since this topic is crucial in setting safety stocks as inadequate estimates of the forecast errors variance can lead to unnecessary inventory holding costs, as well as losses in sales, customer loyalty and revenue, we first discussed the correlation of forecast errors over lead times. We reviewed a commonly employed approximations in practice, and highlighted the analytical drawbacks associated with their use. We then suggested to measure instead the empirical variance of the forecast errors over lead time, and compared it with the standard approximation under different uncertainty settings. Unlike the other approximations, our empirical estimate accounts for the aforementioned correlations, themselves a result of the recursive estimation of multiple-steps-ahead forecasts that are required for lead time forecasting. Our findings indicated that the proposed approximation outperformed the conventional one in terms of inventory performance, especially in the case of forecasting model uncertainty, which is the norm in practice, and then validated our results with real data.

After having determined an adequate method for estimating forecasting uncertainty over lead time, we next turned our attention to capturing it in the case of a supply chain experiencing the Bullwhip Effect in Chapter 3. As forecasting uncertainty has been established as one of contributors to the phenomenon, we sought to better understand its impact on the upstream echelons of the supply chain. We first studied the measurement of the Bullwhip Effect, the Bullwhip Ratio (BWR), which is a ratio of variability of upstream to downstream demand. We explained the difference between demand variability and uncertainty, stressing that the latter is the cost driver for supply chain members. We then highlighted the limitations of the currently adopted measure, and proposed a new metric, the Ratio of Forecast Uncertainty (RFU), to capture the upstream propagation of forecast uncertainty, building on the results from Chapter 2. This measure looks at the ratio of lead time forecast uncertainty of the upstream member's demand to the downstream's, and it circumvents some of the limitations of the BWR which are encountered in practice. In order to compare the effectiveness of our measure with BWR, we devised a simulation experiment and assessed the relationship of each metric with the upstream member's inventory costs. Our results indicate that the proposed metrics bears a stronger link to costs than the traditional BWR, making it a useful indicator for improving the inventory performance at the upstream stages of the supply chain, in addition to detecting the upstream propagation of forecast uncertainty. For example, the RFU metric can be used to determine the upstream changes in forecast uncertainty resulting from a lead time reduction initiative from either the upstream or downstream member, as well as getting a view on the changes in upstream inventory costs.

Having established a new supply chain forecasting metric in Chapter 3, we investigated a commonly advocated remedy in the literature to dampen the Bullwhip Effect in Chapter 4, which consists of providing upstream members with the downstream demand information. As there exists a disparity in the modelling approach for information sharing, as well as mixed results regarding this exchange, we employed the RFU to examine the impact of different information sharing strategies on upstream forecasting accuracy, thus revisiting the debate on whether demand information should be shared between the members, and contrast the performance of each approach with regards to inventory costs. We ran a simulation to conduct this comparison, and introduced forecasting uncertainty at all echelons in the supply chain, as well as managerial order adjustments to render a more realistic simulation. We found that relying exclusively on downstream demand to produce forecasts returns the best results on both accuracy and inventory costs, and determined that using both upstream and downstream demand offers preferable results in both performance aspects over no information sharing. Furthermore, within the multivariate information sharing strategies, we found the inclusion of short-term forecasts for customer demand in producing order forecasts to offer cost and forecast accuracy improvements over their exclusion, as they capture relevant short-term downstream demand information that is absent in the longer lags used by the forecasting models. Finally, the presence of order adjustments at the retailer level, which will increase the variability of the upstream incoming orders, does not alter the rankings of the competing models, which is in part due to the use of the Order-Up-To rule. Different replenishment policies might shed light on the impact of information sharing on attenuating the upstream impact of these adjustments.

5.1 Managerial Implications

The results obtained throughout this thesis are of practical relevance for academics and practitioners alike, as they tackle commonly faced issues in supply chain management. Indeed, as the Bullwhip Effect is a real problem encountered in various industries (for e.g., Bray and Mendelson, 2012; Zotteri, 2013; Shan et al., 2014), the findings from each chapter can be applied in several contexts. For instance, the approximation for lead time forecast errors derived in Chapter 2 is a versatile heuristic as it is independent of any demand assumptions, and can thus be used irrespective of the demand process. In addition, it can be easily implemented within any Entreprise Resource Planning (ERP) or Forecasting Support System

(FSS) software. It can greatly benefit managers in better safety stock levels, which in turn enhance inventory performance and lower deviations from the target service level.

For upstream supply chain members deciding on whether to embark on any information exchange agreement with their downstream partners, the results from Chapters 3 and 4 can assist them in making this decision. The RFU measure, derived in Chapter 3, is actionable as it helps upstream members assess the propagation of forecast uncertainty at the current stage in the supply chain, which allows them to get a clearer view on the impact of improving the forecasting accuracy either at their level or at the retailer level, as well as gauge whether forecast uncertainty is amplifying upstream as a result of the Bullwhip Effect. This ratio can be used, for instance, in determining whether a VMI system should be implemented, as the forecasts would also be centralised. In addition, since the RFU is a forecast accuracy metric, it can be employed to assess the potential improvements in obtaining POS data on the upstream forecasting accuracy. Given its relationship to inventory costs, managers can get a feel of the impact of improving their forecasting accuracy on inventory costs. One drawback related to the use of the RFU by the upstream entities is that it is contingent on the demand and lead time information of both their downstream partner and theirs.

The results in Chapter 4 demonstrate the benefits of information sharing for upstream members in the supply chain. We conclude that POS data improves the forecasting accuracy and amounts to inventory cost reductions, and can thus persuade supply chain members to engage in this information exchange scheme or similar ones such as Quick Response practices, CPFR or VMI, in order to enhance the operations of the supply chain both locally and globally.

5.2 Future Research

The results derived in this thesis have shown the impact of forecast uncertainty on the inventory performance of a supply chain facing the Bullwhip Effect. However, some assumptions were imposed in the experimental setup, and relaxing them could shed additional insights on the value of our findings.

For instance, the lead times were assumed to be deterministic in all three chapters. In practice, the lead times are not fixed, but stochastic. The variability in lead time is accounted for in the total variance estimation in safety stocks (Eppen and Martin, 1988), and thus extending the design from Chapter 2 to incorporate this additional variability would be beneficial in better understanding how to capture the demand uncertainty component for orders and inventory purposes. Similarly, throughout the simulation, the Order-Up-To inventory policy was employed, and given that itself has been identified as a cause for the Bullwhip Effect (Disney and Towill, 2003b), addressing the research questions posed in this thesis with other inventory policies would be an interesting topic to explore, as each impacts the Bullwhip Effect differently (Pillai et al., 2014). This is especially true for our findings in Chapter 4, where the introduction of managerial adjustments to orders did not affect the rankings of our results.

We studied a dyadic supply chain in all three chapters to answer our research questions. Various supply chain structures affect the propagation of order variability in different ways (Hwarng et al., 2005; Giard and Sali, 2013), and thus looking at more complex supply chain setups could provide further insights on our findings. With the exception of Chapter 2, no empirical validation was conducted to substantiate our results, given the difficulty in collecting real supply chain data. An empirical assessment of our findings via a case study will further reinforce the strength and relevance of the results obtained in this thesis.

One notable extension of our work would be to address the topic of a supply chain facing the Bullwhip Effect when promotions are present. Retailer promotions have been identified as one of the four causes of the Bullwhip Effect and is closely related to the demand signal processing cause of the phenomenon studied throughout this thesis. Promotions lead to swings in the demand at the retailer level, which carry on upstream (Lee et al., 1997a,b). Despite promotions being used by many retailers in real life for various purposes, this topic has received limited coverage in the Bullwhip literature, with notable exceptions found in the studies by O'Donnell et al. (2006, 2009); Zhang and Burke (2011); Su and Geunes (2012); Trapero and Pedregal (2016). Many directions for future work remain within this area, as comparing the consequences of the Bullwhip Effect due to promotional campaigns can reflect their upstream impact.

Bibliography

Ackoff, R. L., 1979. The future of operational research is past. Journal of the operational research society 30 (2), 93–104.

Agrawal, S., Sengupta, R. N., Shanker, K., 2009. Impact of information sharing and lead time on Bullwhip Effect and on-hand inventory. European Journal of Operational Research 192 (2), 576–593.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Akkermans, H., Voss, C., 2013. The service Bullwhip Effect. International Journal of Operations & Production Management 33 (6), 765–788.

Ali, M. M., Babai, M. Z., Boylan, J. E., Syntetos, A. A., 2017. Supply chain forecasting when information is not shared. European Journal of Operational Research 260 (3), 984–994.

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

Ali, M. M., Boylan, J. E., 2012. On the effect of non-optimal forecasting methods on supply chain downstream demand. IMA Journal of Management Mathematics 23 (1), 81–98.

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

Ansley, C. F., Newbold, P., 1981. On the bias in estimates of forecast Mean Squared Error. Journal of the American Statistical Association 76 (375), 569–578.

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., 2003. A time-series framework for supply-chain inventory management. Operations Research 51 (2), 210–227.

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

Axsäter, S., 2015. Inventory Control, 3rd Edition. International Series in Operations Research & Management Science. Springer.

Babai, M. Z., Ali, M. M., Boylan, J. E., Syntetos, A. A., 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., Boylan, J. E., Syntetos, A. A., Ali, M. M., 2016. Reduction of the value of information sharing as demand becomes strongly auto-correlated. International Journal of Production Economics 181, 130 – 135.

Badinelli, R. D., 1990. The inventory costs of common mis-specification of demand-forecasting models. International Journal of Production Research 28 (12), 2321–2340.

Balakrishnan, A., Geunes, J., Pangburn, M. S., 2004. Coordinating supply chains by controlling upstream variability propagation. Manufacturing & Service Operations Management 6 (2), 163–183.

Bansal, S., Moritz, B., 2015. Perceived versus actual value of product substitution flexibility: An experimental investigation. Journal of Operations Management 38, 56–70.

Barrow, D. K., Kourentzes, N., 2016. Distributions of forecasting error of forecast combinations: Implications for inventory management. International Journal of Production Economics 177, 24–33.

Bertrand, J. W. M., Fransoo, J. C., 2002. Operations management research methodologies using quantitative modeling. International Journal of Operations & Production Management.

Bhattacharya, R., Bandyopadhyay, S., 2011. A review of the causes of Bullwhip Effect in a supply chain. The International Journal of Advanced Manufacturing Technology 54, 1245–1261.

Bloomfield, R. J., Kulp, S. L., 2013. Durability, transit lags, and optimality of inventory management decisions. Production and Operations Management 22 (4), 826–842.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. Time Series Analysis, Forecasting and Control, 5th Edition. Wiley-Blackwell.

Boylan, J. E., Babai, M. Z., 2016. On the performance of overlapping and non-overlapping temporal demand aggregation approaches. International Journal of Production Economics 181 Part A, 136–144.

Bray, R. L., Mendelson, H., 2012. Information transmission and the Bullwhip Effect: an empirical investigation. Management Science 58 (5), 860–875.

Bray, R. L., Mendelson, H., 2015. Production smoothing and the Bullwhip Effect. Manufacturing & Service Operations Management 17 (2), 208–220.

Breiman, L., 2001. Statistical modeling: The two cultures. Statistical Science 16 (3), 199–231.

Briggs, W., 2016. Uncertainty: The Soul of Modeling, Probability & Statistics. Springer.

Burnham, K. P., Anderson, D. R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 1st Edition. Springer.

Cachon, G. P., 1999. Managing supply chain demand variability with scheduled ordering policies. Management Science 45 (6), 843–856.

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

Cachon, G. P., Lariviere, M. A., 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. Management Science 47 (5), 629–646.

Cachon, G. P., Randall, T., Schmidt, G. M., Fall 2007. In search of the Bullwhip Effect. Manufacturing & Service Operations Management 9 (4), 457–479.

Cannella, S., 2014. Order-Up-To policies in information exchange supply chains. Applied Mathematical Modelling 38 (23), 5553–5561.

Cannella, S., Barbosa-Póvoa, a. P., Framinan, J. M., Relvas, S., 2013. Metrics for Bullwhip Effect analysis. Journal of the Operational Research Society 64 (1), 1–16.

Cannella, S., Ciancimino, E., Framinan, J. M., 2011. Inventory policies and information sharing in multi-echelon supply chains. Production Planning & Control 22 (7), 649–659.

Cattani, K. D., Jacobs, F. R., Schoenfelder, J., 2011. Common inventory modeling assumptions that fall short: Arborescent networks, poisson demand, and single-echelon approximations. Journal of Operations Management 29 (5), 488–499.

Cavusoglu, H., Cavusoglu, H., Raghunathan, S., 2012. Value of and interaction between production postponement and information sharing strategies for supply chain firms. Production and Operations Management 21 (3), 470–488.

Chatfield, C., 1993. Calculating interval forecasts. Journal of Business & Economic Statistics 11 (2), 121–135.

Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society. Series A (Statistics in Society) 158 (3), 419–466.

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

Chatfield, C., 2000. Time-series forecasting. Chapman and Hall/CRC.

Chatfield, C., Koehler, A. B., 1991. On confusing lead time demand with h-period-ahead forecasts. International Journal of Forecasting 7 (2), 239–240.

Chatfield, D. C., Kim, J. G., Harrison, T. P., Hayya, J. C., 2004. The Bullwhip Effect: Impact of stochastic lead time, information quality, and information sharing: A simulation study. Production and Operations Management 13 (4), 340–353.

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

Chen, F., Ryan, J. K., Simchi-Levi, D., 2000b. The impact of exponential smoothing forecasts on the Bullwhip Effect. Naval Research Logistics (NRL) 47 (4), 269–286.

Chen, F., Samroengraja, R., 2004. Order volatility and supply chain cost. Operations Research 52 (5), 707–722.

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

Chen, L., Lee, H. L., July-August 2012. Bullwhip Effect measurement and its implications. Operations Research 60 (4), 771–784.

Chen, L., Luo, W., Shang, K., 2017. Measuring the Bullwhip Effect: Discrepancy and alignment between information and material flows. Manufacturing & Service Operations Management 19 (1), 36–51.

Cheng, T., Wu, Y., 2005. The impact of information sharing in a two-level supply chain with multiple retailers. Journal of the Operational Research Society 56 (10), 1159–1165.

Chevillon, G., 2007. Direct multi-step estimation and forecasting. Journal of Economic Surveys 21 (4), 746–785.

Chiang, C.-Y., Lin, W. T., Suresh, N. C., 2016. An empirically-simulated investigation of the impact of demand forecasting on the Bullwhip Effect: Evidence from U.S. auto industry. International Journal of Production Economics 177, 53–65.

Chiasson, M., Fildes, R., Pidd, M., 2006. Intelligent thinking instead of critical realism? Journal of the Operational Research Society 57 (11), 1373–1375.

Cho, D. W., Lee, Y. H., 2012. Bullwhip Effect measure in a seasonal supply chain. Journal of Intelligent Manufacturing 23, 2295–2305.

Cho, D. W., Lee, Y. H., 2013. The value of information sharing in a supply chain with a seasonal demand process. Computers & Industrial Engineering 65 (1), 97–108.

Chong, Y. Y., Hendry, D. F., 1986. Econometric evaluation of linear macro-economic models. The Review of Economic Studies 53 (4), 671–690.

Churchman, C. W., 1994. Management science: Science of managing and managing of science. Interfaces 24 (4), 99–110.

Clarke, K. A., 2005. The phantom menace: omitted variable bias in econometric research. Conflict management and peace science 22 (4), 341–352.

Costantino, F., Di Gravio, G., Shaban, A., Tronci, M., 2015. The impact of information sharing on ordering policies to improve supply chain performances. Computers & Industrial Engineering 82, 127–142.

Croson, R., Donohue, K., 2005. Upstream versus downstream information and its impact on the Bullwhip Effect. System Dynamics Review: The Journal of the System Dynamics Society 21 (3), 249–260.

Croson, R., Donohue, K., 2006. Behavioral causes of the Bullwhip Effect and the observed value of inventory information. Management Science 52 (3), 323–336.

Cui, R., Allon, G., Bassamboo, A., Van Mieghem, J. A., 2015. Information sharing in supply chains: An empirical and theoretical valuation. Management Science 61 (11), 2803–2824.

Cui, R., Shin, H., 2017. Sharing aggregate inventory information with customers: Strategic cross-selling and shortage reduction. Management Science 64 (1), 381–400.

Davydenko, A., Fildes, R., 2013. Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. International Journal of Forecasting 29 (3), 510 – 522.

Dejonckheere, J., Disney, S., Lambrecht, M., Towill, D., 2003. Measuring and avoiding the Bullwhip Effect: a control theoretic approach. European Journal of Operational Research 147, 567–590.

Disney, S., Towill, D., 2003a. The effect of Vendor Managed Inventory (vmi) dynamics on the Bullwhip Effect in supply chains. International Journal of Production Economics 85, 199–215.

Disney, S. M., Towill, D. R., 2002. A discrete transfer function model to determine the dynamic stability of a Vendor Managed Inventory supply chain. International Journal of Production Research 40 (1), 179–204.

Disney, S. M., Towill, D. R., 2003b. On the Bullwhip and inventory variance produced by an ordering policy. Omega 31, 157–167.

Dong, L., Lee, H. L., 2003. Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand. Operations Research 51 (6), 969–980.

Duan, Y., Yao, Y., Huo, J., 2015. Bullwhip Effect under substitute products. Journal of Operations Management 36, 75–89.

Eppen, G. D., Martin, R., 1988. Determining safety stock in the presence of stochastic lead time and demand. Management Science 34 (11), 1380–1390.

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

Feiler, D. C., Tong, J. D., Larrick, R. P., 2013. Biased judgment in censored environments. Management Science 59 (3), 573–591.

Fildes, R., 1985. Quantitative forecasting—the state of the art: econometric models. Journal of the Operational Research Society 36 (7), 549–580.

Fildes, R., 1992. The evaluation of extrapolative forecasting methods. International Journal of Forecasting 8 (1), 81–98.

Fildes, R., Goodwin, P., 2007. Against your better judgment? How organizations can improve their use of management judgment in forecasting. Interfaces 37 (6), 570–576.

Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos, K., 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., Kingsman, B., 2011. Incorporating demand uncertainty and forecast error in supply chain planning models. Journal of the Operational Research Society 62, 1–18.

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

Fleischhacker, A. J., Fok, P.-W., 2015. On the relationship between entropy, demand uncertainty, and expected loss. European Journal of Operational Research 245 (2), 623–628.

Fleming, P. J., Wallace, J. J., 1986. How not to lie with statistics: The correct way to summarize benchmark results. Communications of the ACM 29 (3), 218–221.

Flores, B., Olson, D., Pearce, S., 1993. Use of cost and accuracy measures in forecasting method selection: a physical distribution example. International Journal of Production Research 31 (1), 139–160.

Flynn, B. B., Sakakibara, S., Schroeder, R. G., Bates, K. A., Flynn, E., 1990. Empirical research methods in operations management. Journal of Operations Management 9 (2), 250 – 284.

Forrester, J., 1961. Industrial Dynamics. MIT University Press.

Fransoo, J. C., Wouters, M. J., 2000. Measuring the Bullwhip Effect in the supply chain. Supply Chain Management: An International Journal 5 (2), 78–89.

Fricker, R. D., Goodhart, C. A., 2000. Applying a bootstrap approach for setting reorder points in military supply systems. Naval Research Logistics 47 (6), 459–478.

Gaalman, G., Disney, S. M., 2006. State space investigation of the Bullwhip problem with ARMA(1,1) demand processes. International Journal of Production Economics 104, 327–339.

Gaalman, G., Disney, S. M., 2009. On Bullwhip in a family of Order-Up-To policies with ARMA(2,2) demand and arbitrary lead-times. International Journal of Production Economics 121 (2), 454–463.

Ganesh, M., Raghunathan, S., Rajendran, C., 2014. The value of information sharing in a multi-product, multi-level supply chain: Impact of product substitution, demand correlation, and partial information sharing. Decision Support Systems 58, 79–94.

Gardner, E. S., 1988. A simple method of computing prediction intervals for time series forecasting. Management Science 34 (4), 541–546.

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

Gardner Jr, E. S., 2006. Exponential smoothing: The state of the art - part ii. International journal of forecasting 22 (4), 637–666.

Gaur, V., Fisher, M. L., Raman, A., 2005a. An econometric analysis of inventory turnover performance in retail services. Management Science 51 (2), 181–194.

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

Ghysels, E., Osborn, D. R., 2001. The econometric analysis of seasonal time series. Themes in Modern Econometrics. Cambridge University Press.

Giard, V., Sali, M., 2013. The Bullwhip Effect in supply chains: a study of contingent and incomplete literature. International Journal of Production Research 51 (13), 3880–3893.

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

Giloni, A., Hurvich, C., Seshadri, S., 2014. Forecasting and information sharing in supply chains under ARMA demand. IIE Transactions 46 (1), 35–54.

Gneiting, T., 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106 (494), 746–762.

Granger, C. W., Lee, T.-H., 1989. Investigation of production, sales and inventory relationships using multicointegration and non-symmetric error correction models. Journal of Applied Econometrics 4 (S1), S145–S159.

Graves, S. C., 1999. A single-item inventory model for a nonstationary demand process. Manufacturing & Service Operations Management 1 (1), 50–61.

Hammond, J. H., 1994. Barilla SpA (A). Harvard Business School Boston, MA.

Hançerlioğulları, G., Şen, A., Aktunç, E. A., 2016. Demand uncertainty and inventory turnover performance: An empirical analysis of the US retail industry. International Journal of Physical Distribution & Logistics Management 46 (6/7), 681–708.

Hanssens, D. M., 1998. Order forecasts, retail sales, and the marketing mix for consumer durables. Journal of Forecasting 17 (3-4), 327–346.

Harrison, J. R., Lin, Z., Carroll, G. R., Carley, K. M., 2007. Simulation modeling in organizational and management research. Academy of Management Review 32 (4), 1229–1245.

Hartzel, K. S., Wood, C. A., 2017. Factors that affect the improvement of demand forecast accuracy through point-of-sale reporting. European Journal of Operational Research 260 (1), 171–182.

Haughton, M. A., 2009. Distortion Bullwhip Effect on carriers. Transportation Research Part E 45, 172–185.

Hausman, W. H., 1969. Sequential decision problems: A model to exploit existing forecasters. Management Science 16 (2), B93–B111.

Heath, D. C., Jackson, P., 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Transactions 26 (3), 17–30.

Hendry, D. F., Learmer, E. E., Poirier, D. J., 1990. A conversation on econometric methodology. Econometric Theory 6 (2), 171–261.

Hollander, M., Wolfe, D. A., Chicken, E., 2013. Nonparametric statistical methods. John Wiley & Sons.

Holst, J., Holst, U., Madsen, H., Melgaard, H., 1993. Validation of grey box models. In: Adaptive Systems in Control and Signal Processing 1992. Elsevier, pp. 53–60.

Holweg, M., Disney, S., Holmström, J., Småros, J., 2005. Supply chain collaboration: Making sense of the strategy continuum. European Management Journal 23 (2), 170–181.

Hosoda, T., Disney, S., 2009. Impact of market demand mis-specification on a two-level supply chain. International Journal of Production Economics 121, 739–751.

Hosoda, T., Disney, S. M., 2006. On variance amplification in a three-echelon supply chain with Minimum Mean Square Error forecasting. Omega 34, 344–358.

Hosoda, T., Naim, M. M., Disney, S. M., Potter, A., 2008. Is there a benefit to sharing market sales information? Linking theory and practice. Computers & Industrial Engineering 54 (2), 315–326.

Huang, Y.-S., Hung, J.-S., Ho, J.-W., 2017. A study on information sharing for supply chains with multiple suppliers. Computers & Industrial Engineering 104, 114–123.

Hwarng, H., Chong, C., Xie, N., Burgess, T., 2005. Modelling a complex supply chain: understanding the effect of simplified assumptions. International Journal of Production Research 43 (13), 2829–2872.

Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 26 (3), 1–22.

Hyndman, R., Koehler, A., Ord, K., Snyder, R., 2008. Forecasting with Exponential Smoothing: The State Space Approach. Springer.

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

Isaksson, O. H., Seifert, R. W., 2016. Quantifying the Bullwhip Effect using two-echelon data: A cross-industry empirical investigation. International Journal of Production Economics 171, 311–320.

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

Jakšič, M., Rusjan, B., 2008. The effect of replenishment policies on the Bullwhip Effect: A transfer function approach. European Journal of Operational Research 184 (3), 946 – 961.

Janssen, E., Strijbosch, L. W., Brekelmans, R., 2009. Assessing the effects of using demand parameters estimates in inventory control and improving the performance using a correction function. International Journal of Production Economics 118 (1), 34–42.

Jin, M., DeHoratius, N., Schmidt, G., 2017. In search of intra-industry bullwhips. International Journal of Production Economics 191, 51–65.

Jin, Y., Williams, B. D., Tokar, T., Waller, M. A., 2015a. Forecasting with temporally aggregated demand signals in a retail supply chain. Journal of Business Logistics 36 (2), 199–211.

Jin, Y., Williams, B. D., Waller, M. A., Hofer, A. R., 2015b. Masking the Bullwhip Effect in retail: the influence of data aggregation. International Journal of Physical Distribution & Logistics Management 45 (8), 814–830.

Johnston, F. R., Harrison, P. J., 1986. The variance of lead time demand. Journal of the Operational Research Society 37 (3), 303–308.

Jurado, K., Ludvigson, S. C., Ng, S., March 2015. Measuring uncertainty. The American Economic Review 105 (3), 1177–1216.

Kahn, J., 1987. Inventories and the volatility of production. The American Economic Review 77 (4), 667–679.

Kaipia, R., Holmström, J., Småros, J., Rajala, R., 2017. Information sharing for sales and operations planning: Contextualized solutions and mechanisms. Journal of Operations Management 52, 15–29.

Kelepouris, T., Miliotis, P., Pramatari, K., 2008. The impact of replenishment parameters and information sharing on the Bullwhip Effect: A computational study. Computers & Operations Research 35 (11), 3657–3670.

Kelle, P., Milne, A., 1999. The effect of (s,S) ordering policy on the supply chain. International Journal of Production Economics 59 (1), 113 – 122.

Kembro, J., Näslund, D., 2014. Information sharing in supply chains, myth or reality? A critical analysis of empirical literature. International Journal of Physical Distribution & Logistics Management 44 (3), 179–200.

Kembro, J., Selviaridis, K., 2015. Exploring information sharing in the extended supply chain: an interdependence perspective. Supply Chain Management: An International Journal 20 (4), 455–470.

Kembro, J., Selviaridis, K., Näslund, D., 2014. Theoretical perspectives on information sharing in supply chains: a systematic literature review and conceptual framework. Supply chain management: An international journal 19 (5/6), 609–625.

Kemppainen, K., Vepsäläinen, A. P., 2003. Trends in industrial supply chains and networks. International journal of physical distribution & logistics management 33 (8), 701–719.

Kerkkänen, A., Korpela, J., Huiskonen, J., 2009. Demand forecasting errors in industrial context: Measurement and impacts. International Journal of Production Economics 118, 43–48.

Kesavan, S., Gaur, V., Raman, A., 2010. Do inventory and gross margin data improve sales forecasts for US public retailers? Management Science 56 (9), 1519–1533.

Kim, H., Ryan, J. K., 2003. The cost impact of using simple forecasting techniques in a supply chain. Naval Research Logistics 50, 388–411.

Kim, J. G., Chatfield, D. C., Harrison, T. P., Hayya, J. C., 2006. Quantifying the Bullwhip Effect in a supply chain with stochastic lead time. European Journal of Operational Research 173 (2006), 617–636.

Kleijnen, J. P., 1995. Verification and validation of simulation models. European journal of operational research 82 (1), 145–162.

Koehler, A. B., 1990. An inappropriate prediction interval. International Journal of Forecasting 6 (4), 557–558.

Kourentzes, N., 2013. Intermittent demand forecasts with Neural Networks. International Journal of Production Economics 143 (1), 198–206.

Kourentzes, N., Athanasopoulos, G., 2019. Cross-temporal coherent forecasts for Australian tourism. Annals of Tourism Research 75, 393–409.

Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting 30 (2), 291–302.

Kourentzes, N., Trapero, J. R., Barrow, D. K., 2020. Optimising forecasting models for inventory planning. International Journal of Production Economics 225, 107597.

Kovtun, V., Giloni, A., Hurvich, C., 2019. The value of sharing disaggregated information in supply chains. European Journal of Operational Research 277 (2), 469–478.

Kremer, M., Minner, S., Van Wassenhove, L. N., 2010. Do random errors explain newsvendor behavior? Manufacturing & Service Operations Management 12 (4), 673–681.

Kulp, S. C., Lee, H. L., Ofek, E., 2004. Manufacturer benefits from information integration with retail customers. Management Science 50 (4), 431–444.

Kunnumkal, S., Topaloglu, H., 2008. Price discounts in exchange for reduced customer demand variability and applications to advance demand information acquisition. International Journal of Production Economics 111 (2), 543 – 561.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics 54 (1), 159 – 178.

Law, A., 2014. Simulation Modeling and Analysis, 5th Edition. McGraw-Hill Education.

Lee, H., Whang, S., 1999. Decentralized multi-echelon supply chains: Incentives and information. Management Science 45 (5), 633–640.

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

Lee, H. L., Padmanabhan, V., Whang, S., 1997b. Information distortion in a supply chain: the Bullwhip Effect. Management Science 43 (4), 546–558.

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

Lee, Y. S., 2014. A semi-parametric approach for estimating critical fractiles under autocorrelated demand. European Journal of Operational Research 234, 163–173.

Lee, Y. S., Scholtes, S., 2014. Empirical prediction intervals revisited. International Journal of Forecasting 30 (2), 217–234.

Li, G., Wang, S., Yan, H., Yu, G., 2005. Information transformation in a supply chain: a simulation study. Computers & Operations Research 32 (3), 707 – 725.

Li, L., 2002. Information sharing in a supply chain with horizontal competition. Management Science 48 (9), 1196–1212.

Li, S., Cui, X., Huo, B., Zhao, X., 2019. Information sharing, coordination and supply chain performance: The moderating effect of demand uncertainty. Industrial Management & Data Systems 19 (5), 1046–1071.

Liao, T. W., Chang, P., 2010. Impact of forecast, inventory policy, and lead time on supply chain inventory–a numerical study. International Journal of Production Economics 128, 527–537.

Luong, H. T., Phien, N. H., 2007. Measure of Bullwhip Effect in supply chains: The case of high order autoregressive demand process. European Journal of Operational Research 183 (1), 197–209.

Ma, J., Ma, X., 2017. Measure of the Bullwhip Effect considering the market competition between two retailers. International Journal of Production Research 55 (2), 313–326.

Ma, S., Fildes, R., Huang, T., 2016. Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. European Journal of Operational Research 249 (1), 245–257.

Makridakis, S., Winkler, R. L., 1989. Sampling distributions of post-sample forecasting errors. Journal of the Royal Statistical Society. Series C (Applied Statistics), 38 (2), 331–342.

Manary, M. P., Willems, S. P., 2008. Setting safety-stocks targets at Intel in the presence of forecast bias. Interfaces 38 (2), 112–122.

Manary, M. P., Willems, S. P., Shihata, A. F., 2009. Correcting heterogeneous and biased forecast error at Intel for supply chain optimization. Interfaces 39 (5), 415–427.

McElroy, T., 2015. When are direct multi-step and iterative forecasts identical? Journal of Forecasting 34 (4), 315–336.

Meredith, J. R., 2001. Reconsidering the philosophical basis of OR/MS. Operations Research 49 (3), 325–333.

Metters, R., 1997. Quantifying the Bullwhip Effect in supply chains. Journal of Operations Management 15, 89–100.

Michna, Z., Nielsen, P., Nielsen, I. E., 2018. The impact of stochastic lead times on the Bullwhip Effect: a theoretical insight. Production & Manufacturing Research 6 (1), 190–200.

Mingers, J., 2003. A classification of the philosophical assumptions of management science methods. Journal of the operational research society 54 (6), 559–570.

Mingers, J., 2006. A critique of statistical modelling in management science from a critical realist perspective: its role within multimethodology. Journal of the Operational Research Society 57 (2), 202–219.

Miragliotta, G., 2006. Layers and mechanisms: A new taxonomy for the Bullwhip Effect. International Journal of production economics 104 (2), 365–381.

Mishra, B. K., Raghunathan, S., Yue, X., 2007. Information sharing in supply chains: Incentives for information distortion. IIE Transactions 39 (9), 863–877.

Mishra, B. K., Raghunathan, S., Yue, X., 2009. Demand forecast sharing in supply chains. Production and Operations Management 18 (2), 152–166.

Nagaraja, C., Thavaneswaran, A., Appadoo, S., 2015. Measuring the Bullwhip Effect for supply chains with seasonal demand components. European Journal of Operational Research 242 (2), 445–454.

Nagaraja, C. H., McElroy, T., 2018. The multivariate Bullwhip Effect. European Journal of Operational Research 267, 96–106.

Naylor, T. H., Finger, J. M., 1967. Verification of computer simulation models. Management science 14 (2), B–92.

Nepal, B., Murat, A., Chinnam, R. B., 2012. The Bullwhip Effect in capacitated supply chains with consideration for product life-cycle-aspects. International Journal of Production Economics 136, 318–331.

Nielsen, E. H., 2013. Small sample uncertainty aspects in relation to Bullwhip Effect measurement. International Journal of Production Economics 146, 543–549.

Niranjan, T. T., Wagner, S. M., Bode, C., 2011. An alternative theoretical explanation and empirical insights into overordering behavior in supply chains. Decision Sciences 42 (4), 859–888.

Ord, K., Fildes, R., Kourentzes, N., 2017. Principles of business forecasting, 2nd Edition. Wessex Press Publishing Co.

Ouyang, Y., 2007. The effect of information sharing on supply chain stability and the Bullwhip Effect. European Journal of Operational Research 182 (3), 1107–1121.

Ouyang, Y., Daganzo, C., 2006. Characterization of the Bullwhip Effect in linear, time-invariant supply chains. Management Science 52 (10), 1544–1556.

Ouyang, Y., Daganzo, C., 2008. Robust tests for the Bullwhip Effect in supply chains with stochastic dynamics. European Journal of Operational Research 185, 340–353.

O'Donnell, T., Humphreys, P., McIvor, R., Maguire, L., 2009. Reducing the negative effect of sales promotions in supply chains using genetic algorithms. Expert Systems with Applications 36 (4), 7827–7837.

O'Donnell, T., Maguire, L., McIvor, R., Humphreys, P., 2006. Minimizing the Bullwhip Effect in a supply chain using genetic algorithms. International Journal of Production Research 44 (8), 1523–1543.

Papier, F., 2016. Supply allocation under sequential advance demand information. Operations Research 64 (2), 341–361.

Pastore, E., Alfieri, A., Zotteri, G., 2019a. An empirical investigation on the antecedents of the Bullwhip Effect: Evidence from the spare parts industry. International Journal of Production Economics 209, 121 – 133.

Pastore, E., Alfieri, A., Zotteri, G., Boylan, J. E., 2019b. The impact of demand parameter uncertainty on the Bullwhip Effect. European Journal of Operational Research.

Phillips, P. C., 1979. The sampling distribution of forecasts from a first-order autoregression. Journal of Econometrics 9, 241–261.

Pidd, M., 2009. Tools for Thinking: Modelling in Management Science, 3rd Edition. Wiley.

Pillai, M., Talari, P. C., Elluri, P. V., 2014. Performance analysis of some supply chain replenishment strategies. International Journal of Logistics Research and Applications 17 (5), 357–376.

Prak, D., Teunter, R., Syntetos, A., 2017. On the calculation of safety stocks when demand is forecasted. European Journal of Operational Research 256, 454–461.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

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

Raghunathan, S., 2003. Impact of demand correlation on the value of and incentives for information sharing in a supply chain. European Journal of Operational Research 146 (3), 634–649.

Ray, W. D., 1982. ARIMA forecasting models in inventory control. Journal of the Operational Research Society 33 (6), 567–574.

Reiner, G., Fichtinger, J., 2009. Demand forecasting for supply processes in consideration of pricing and market information. International Journal of Production Economics 118 (1), 55–62.

Ren, Y., Croson, R., 2013. Overconfidence in newsvendor orders: An experimental study. Management Science 59 (11), 2502–2517.

Ritchken, P. H., Sankar, R., 1984. The effect of estimation risk in establishing safety stock levels in an inventory model. Journal of the Operational Research Society 35 (12), 1091–1099.

Robinson, S., 2014. Simulation: the Practice of Model Development and Use, 2nd Edition. Red Globe Press.

Rostami-Tabar, B., Ali, M. Z. B. M., Boylan, J. E., 2019. The impact of temporal aggregation on supply chains with ARMA(1,1) demand processes. European Journal of Operational Research 273 (3), 920 – 932.

Rumyantsev, S., Netessine, S., 2007. What can be learned from classical inventory models? A cross-industry exploratory investigation. Manufacturing & Service Operations Management 9 (4), 409–429.

Sabitha, D., Rajendran, C., Kalpakam, S., Ziegler, H., 2016. The value of information sharing in a serial supply chain with AR(1) demand and non-zero replenishment lead times. European Journal of Operational Research 255 (3), 758–777.

Sadeghi, A., 2015. Providing a measure for Bullwhip Effect in a two-product supply chain with exponential smoothing forecasts. International Journal of Production Economics 169, 44–54.

Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., Desmet, B., 2018. Tactical sales forecasting using a very large set of macroeconomic indicators. European Journal of Operational Research 264 (2), 558–569.

Sanders, N. R., Graman, G. A., 2009. Quantifying costs of forecast errors: a case study of the warehouse environment. Omega 37 (1), 116–125.

Sanders, N. R., Graman, G. A., 2016. Impact of bias magnification on supply chain costs: The mitigating role of forecast sharing. Decision Sciences 47 (5), 881–906.

Saoud, P., Kourentzes, N., Boylan, J. E., 2018. Approximations for the lead time variance: The omitted sources of uncertainty, Lancaster University, Working Paper.

Saoud, P., Kourentzes, N., Boylan, J. E., 2019. Forecast uncertainty as a determinant for the Bullwhip Effect, Lancaster University, Working Paper.

Sargent, R. G., 2013. Verification and validation of simulation models. Journal of Simulation 7 (1), 12–24.

Schruben, L., Singh, H., Tierney, L., 1983. Optimal tests for initialization bias in simulation output. Operations Research 31 (6), 1167–1178.

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

Shafer, S. M., Smunt, T. L., 2004. Empirical simulation studies in operations management: context, trends, and research opportunities. Journal of Operations Management 22 (4), 345–354.

Shan, J., Yang, S., Yang, S., Zhang, J., April 2014. An empirical study of the Bullwhip Effect in China. Production and Operations Management 23 (4), 537–551.

Shang, W., Ha, A. Y., Tong, S., 2015. Information sharing in a supply chain with a common retailer. Management Science 62 (1), 245–263.

Silver, E. A., 1981. Operations research in inventory management: A review and critique. Operations Research 29 (4), 628–645.

Silver, E. A., Pyke, D. F., Thomas, D. J., 2016. Inventory and Production Management in Supply Chains, 4th Edition. CRC Press.

Silver, E. A., Rahnama, M. R., 1986. The cost effect of statistical sampling in selecting the reorder point in a common inventory model. The Journal of the Operational Research Society 37 (7), 705–713.

Silver, E. A., Rahnama, M. R., 1987. Biased selection of the inventory reorder point when demand parameters are statistically estimated. Engineering Costs and Production Economics 12, 283–292.

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. D., Koehler, A. B., Hyndman, R. J., Ord, J. K., 2004. Exponential Smoothing models: Means and variance for lead-time demand. European Journal of Operational Research 158, 444–455.

Snyder, R. D., Koehler, A. B., Ord, J. K., 2002. Forecasting for inventory control with exponential smoothing. International Journal of Forecasting 18, 5–18.

Srivathsan, S., Kamath, M., 2018. Understanding the value of upstream inventory information sharing in supply chain networks. Applied Mathematical Modelling 54, 393–412.

Steckel, J. H., Gupta, S., Banerji, A., 2004. Supply chain decision making: Will shorter cycle times and shared point-of-sale information necessarily help? Management Science 50 (4), 458–464.

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

Sterman, J. D., Dogan, G., 2015. "I'm not hoarding, I'm just stocking up before the hoarders get here.": Behavioral causes of phantom ordering in supply chains. Journal of Operations Management 39, 6–22.

Stößlein, M., Kanet, J. J., Gorman, M., Minner, S., 2014. Time-phased safety stocks planning and its financial impacts: Empirical evidence based on European econometric data. International Journal of Production Economics 149, 47–55.

Strijbosch, L. W., Syntetos, A. A., Boylan, J. E., Janssen, E., 2011. On the interaction between forecasting and stock control: The case of non-stationary demand. International Journal of Production Economics 133, 470–480.

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

Su, Y., Geunes, J., 2012. Price promotions, operations cost, and profit in a two-stage supply chain. Omega 40 (6), 891–905.

Sucky, E., 2009. The Bullwhip Effect in supply chains–an overestimated problem? International Journal of Production Economics 118, 311–322.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., Nikolopoulos, K., 2016. Supply chain forecasting: Theory, practice, their gap and the future. European Journal of Operational Research 252 (1), 1–26.

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

Syntetos, A. A., Boylan, J. E., Disney, S. M., 2009. Forecasting for inventory planning: a 50-year review. Journal of the Operational Research Society 60 (Supplement 1), S149–S160.

Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International journal of forecasting 16 (4), 437–450.

Taylor, J. W., Bunn, D. W., 1999. A quantile regression approach to generating prediction intervals. Management Science 45 (2), 225–237.

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

Teunter, R. H., Babai, M. Z., Bokhorst, J. A., Syntetos, A. A., 2018. Revisiting the value of information sharing in two-stage supply chains. European Journal of Operational Research 270 (3), 1044 – 1052.

Theil, H., 1966. Applied economic forecasting. North Holland.

Thombs, L. A., Schucany, W. R., 1990. Bootstrap prediction intervals for autoregression. Journal of the American Statistical Association 85 (410).

Thonemann, U. W., 2002. Improving supply-chain performance by sharing advance demand information. European Journal of Operational Research 142 (1), 81–107.

Tliche, Y., Taghipour, A., Canel-Depitre, B., 2019. Downstream Demand Inference in decentralized supply chains. European Journal of Operational Research 274 (1), 65–77.

Tokar, T., Aloysius, J., Williams, B., Waller, M., 2014. Bracing for demand shocks: An experimental investigation. Journal of Operations Management 32 (4), 205–216.

Tokar, T., Aloysius, J. A., Waller, M. A., 6 2012. Supply chain inventory replenishment: The debiasing effect of declarative knowledge. Decision Sciences 43 (3), 525–546.

Tokar, T., Aloysius, J. A., Waller, M. A., Williams, B. D., 2011. Retail promotions and information sharing in the supply chain: a controlled experiment. The International Journal of Logistics Management 22 (1), 5–25.

Towill, D. R., Zhou, L., Disney, S. M., 2007. Reducing the Bullwhip Effect: Looking through the appropriate lens. International Journal of Production Economics 108 (1-2), 444–453.

Trapero, J. R., Cardós, M., Kourentzes, N., 2019a. Empirical safety stock estimation based on Kernel and GARCH models. Omega 84, 199–211.

Trapero, J. R., Cardós, M., Kourentzes, N., 2019b. Quantile forecast optimal combination to enhance safety stock estimation. International Journal of Forecasting 35 (1), 239–250.

Trapero, J. R., García, F. P., Kourentzes, N., et al., 2014. Impact of demand nature on the Bullwhip Effect. bridging the gap between theoretical and empirical research. In: Proceedings of the Seventh International Conference on Management Science and Engineering Management. pp. 1127–1137.

Trapero, J. R., Kourentzes, N., Fildes, R., 2012. Impact of information exchange on supplier forecasting performance. Omega 40, 738–747.

Trapero, J. R., Pedregal, D. J., 2016. A novel time-varying Bullwhip Effect metric: An application to promotional sales. International Journal of Production Economics 182, 465 – 471.

Trapero, J. R., Pedregal, D. J., Fildes, R., Kourentzes, N., 2013. Analysis of judgmental adjustments in the presence of promotions. International Journal of Forecasting 29 (2), 234–243.

Urban, T. L., 2005. A periodic-review model with serially-correlated, inventory-level-dependent demand. International Journal of Production Economics 95 (3), 287–295.

Van Der Zee, D.-J., Van Der Vorst, J. G., 2005. A modeling framework for supply chain simulation: opportunities for improved decision making. Decision sciences 36 (1), 65–95.

Van Donselaar, K. H., Gaur, V., Van Woensel, T., Broekmeulen, R. A., Fransoo, J. C., 2010. Ordering behavior in retail stores and implications for automated replenishment. Management Science 56 (5), 766–784.

van Gigch, J. P., 1989. The potential demise of OR/MS: Consequences of neglecting epistemology. European Journal of Operational Research 42 (3), 268–278.

Wagner, H. M., 2002. And then there were none. Operations Research 50 (1), 217–226.

Wan, X., Sanders, N. R., 2017. The negative impact of product variety: Forecast bias, inventory levels, and the role of vertical integration. International Journal of Production Economics 186, 123 – 131.

Wang, M.-C., Rao, S., 1992. Estimating reorder points and other management science applications by bootstrap procedure. European Journal of Operational Research 56, 332–342.

Wang, X., Disney, S. M., 2016. The Bullwhip Effect: Progress, trends and directions. European Journal of Operational Research 250 (3), 691 – 701.

Weiss, A. A., Andersen, A. P., 1984. Estimating time series models using the relevant forecast evaluation criterion. Journal of the Royal Statistical Society. Series A (General) 147 (3), 484–487.

Williams, B. D., Waller, M. A., 2010. Creating order forecasts: point-of-sale or order history? Journal of Business Logistics 31 (2), 231–251.

Williams, B. D., Waller, M. A., Ahire, S., Ferrier, G. D., 2014. Predicting retailer orders with POS and order data: The inventory balance effect. European Journal of Operational Research 232 (3), 593–600.

Wright, D., Yuan, X., 2008. Mitigating the Bullwhip Effect by ordering policies and forecasting methods. International Journal of Production Economics 113, 587–597.

Yao, Y., Dresner, M., 2008. The inventory value of information sharing, continuous replenishment, and vendor-managed inventory. Transportation Research Part E: Logistics and Transportation Review 44 (3), 361–378.

Yee, S.-T., 2005. Impact analysis of customized demand information sharing on supply chain performance. International Journal of Production Research 43 (16), 3353–3373.

Yu, Z., Yan, H., Cheng, T., 2002. Modelling the benefits of information sharing-based partnerships in a two-level supply chain. Journal of the Operational Research Society 53 (4), 436–446.

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

Zhang, X., 2004b. The impact of forecasting methods on the Bullwhip Effect. International Journal of Production Economics 88 (1), 15 – 27.

Zhang, X., 2007. Inventory control under temporal demand heteroscedasticity. European Journal of Operational Research 182, 127–144.

Zhang, X., Burke, G. J., 2011. Analysis of compound Bullwhip Effect causes. European Journal of Operational Research 210, 514–526.

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

Zhao, X., Xie, J., Leung, J., 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.

Zotteri, G., 2013. An empirical investigation on causes and effects of the Bullwhip-Effect: evidence from the personal care sector. International Journal of Production Economics 143, 489–498.