Aalto University
School of Business

IMPROVING THE BUDGETING OF FRAMEWORK AGREEMENT VOLUMES WITH TIME SERIES MODELS IN GOVERNMENTAL PROCUREMENT

Case Hansel

Master's Thesis
Kolari Saara
Aalto University School of Business
Information and Service Management
Spring 2019

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of Master's thesis

Author Kolari Saara

Title of thesis Improving the Budgeting of Framework Agreement Volumes with Time Series Models in Governmental Procurement

Degree Master of Science in Economics and Business Administration

Degree programme Information and Service Management

Thesis advisor(s) Kemppainen Katariina and Kauppi Katri

Year of approval 2019 **Number of pages** 98 **Language** English

Abstract

Companies still use existing data and predictive analytics too rarely to support their decision making. The use of analytical models has also been neglected in public procurement research. This thesis addresses the utilisation of time series analytics in supporting decision making, improving budgeting, and forecasting public procurement volumes through a case company. The aim is to examine how well time series models can forecast the volumes of governmental procurement, and whether these models can improve the budgeting accuracy of the case company.

Hansel Ltd is a non-profit central purchasing body of the Finnish central government whose purpose is to save public funds by increasing the productivity of governmental procurement. Hansel's operations focus on tendering and maintaining framework agreements in accordance with the national and the EU legislation. Since around 87 percent of Hansel's revenue is comprised of framework agreement specific service fees, their allocation is a strategically important decision for achieving zero profit. The level of service fees is based on the budgeted framework agreement volumes, which unfortunately differ from the actual volumes. In recent years, the overall budget has differed from the actual total on average 5 percent, but the differences between the budgeted and the actual volumes vary from few to even 100 percent at the framework agreement level.

The ability of time series models to forecast future framework agreement volumes and the effects of the models on Hansel's budgeting process was studied by modelling 26 framework agreement subtotals. ARIMA models, which predict future values through previous values and forecast errors, were used in the modelling. ARIMA models capture efficiently the temporal dependence of values with a finite number of parameters, making modelling and creating accurate forecasts straightforward. The forecast accuracy of the models was compared to the current budget numerically with the root mean squared error and the mean absolute percentage error as well as graphically.

The forecasts of the models were about 6 percentage points more accurate than the current budget. Also, the differences to the actual volumes were decreased on average 37 percent at the framework agreement subtotal level. In addition, implementing the models as a part of the budgeting process would reduce the number of steps and overlapping work in the process and increase the transparency of budgeting when the budget was based on theory rather than on the intuition of the top management. By increasing the budgeting accuracy and enhancing the budgeting process the models would also improve the allocation of adequate service fees and achieving the zero profit objective. As the ARIMA models were found to be competent to forecast framework agreement volumes accurately, this thesis has practical implications also outside Hansel. Framework agreements have established a key role in European public procurement, and the benefits of utilizing time series models for public procurement units other than Hansel should be further studied.

Keywords public procurement, centralised procurement, time series analytics, ARIMA model, demand forecasting, framework agreement, budgeting

Aalto-yliopisto
Kauppakorkeakoulu

Aalto-yliopisto, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Maisterintutkinnon tutkielman tiivistelmä

Tekijä Kolari Saara

Työn nimi Improving the Budgeting of Framework Agreement Volumes with Time Series Models in Governmental Procurement

Tutkinto Kauppatieteiden maisteri

Koulutusohjelma Information and Service Management

Työn ohjaaja(t) Kemppainen Katariina ja Kauppi Katri

Hyväksymisvuosi 2019 **Sivumäärä** 98 **Kieli** englanti

Tiivistelmä

Yritykset käyttävät olemassa olevaa dataa ja ennustavaa analytiikkaa edelleen liian harvoin päätöksentekonsa tukena. Analyyttisten mallien hyödyntäminen on myös jäänyt vähälle huomiolle julkisten hankintojen tutkimuksessa. Tämä tutkimus käsittelee aikasarja-analytiikan hyödyntämistä päätöksenteon tukemisessa, budjetoinnin kehittämisessä ja julkisten hankintojen volyymin ennustamisessa esimerkkiyrityksen kautta. Tavoitteena on selvittää, kuinka hyvin aikasarjamalleilla kyetään ennustamaan valtion yhteishankintojen volyymeja ja pystytäänkö esimerkkiyrityksen budjetoinnin tarkkuutta parantamaan kyseisten mallien avulla.

Hansel Oy on voittoa tavoittelematon valtion yhteishankintayksikkö, jonka tavoitteena on säästää yhteiskunnan varoja tehostamalla valtion hankintatointa. Hanselin toiminta keskittyy puitejärjestelyjen kilpailuttamiseen ja ylläpitämiseen kansallisen ja EU-lainsäädännön mukaisesti. Koska noin 87 prosenttia Hanselin liikevaihdosta tulee puitejärjestelykohtaisista palvelumaksuista, niiden allokointi on strategisesti tärkeä päätös nollatavoitteen kannalta. Palvelumaksujen suuruus riippuu budjetoiduista puitejärjestelyvolyymeista, jotka eivät valitettavasti vastaa toteutuneita volyymeja. Kokonaisbudjetti on viime vuosina eronnut toteutuneista volyymeistä keskimäärin 5 prosenttia, mutta budjetoitujen ja toteutuneiden volyymien erot vaihtelevat puitejärjestelytasolla muutamasta prosentista jopa 100 prosenttiin.

Aikasarjamallien kyvykkyyttä ennustaa tulevia puitejärjestelyvolyymeja ja vaikutuksia Hanselin budjetointiprosessiin tutkittiin mallintamalla 26 puitejärjestelyvälisummaa. Mallinnuksessa käytettiin ARIMA-malleja, jotka ennustavat tulevia arvoja edellisten arvojen ja ennusteista syntyvien virhetermien kautta. ARIMA-mallit kuvaavat tehokkaasti arvojen ajallista riippuvuutta toisistaan rajallisella määrällä parametreja, mikä tekee mallintamisesta ja tarkkojen ennusteiden luomisesta mutkatonta. Mallien ennustetarkkuutta verrattiin nykyiseen budjettiin sekä numeerisesti keskimääräisellä neliövirheellä ja keskimääräisellä absoluuttisella prosenttivirheellä että graafisesti.

Mallien ennusteet olivat noin 6 prosenttiyksikköä tarkempia kuin nykyinen budjetti, ja välisummatasoiset erot toteutuneisiin volyymeihin laskivat keskimäärin 37 prosenttia. Lisäksi mallien implementointi osaksi budjetointiprosessia vähentäisi prosessin vaiheiden ja päällekkäisen työn määrää ja lisäisi budjetoinnin läpinäkyvyyttä, kun budjetti perustuisi teoriaan sen sijaan, että se perustuu ylimmän johdon intuitioon. Parantamalla budjetoinnin tarkkuutta ja tehostamalla budjetointiprosessia mallit myös edistäisivät oikeansuuruisten palvelumaksujen allokointia ja nollatavoitteen saavuttamista. Koska ARIMA-malleilla pystyttiin ennustamaan puitejärjestelyvolyymeja tarkasti, tutkimuksen tuloksilla on käytännön merkitystä myös Hanselin ulkopuolella. Puitejärjestelyt ovat vakiinnuttaneet asemansa eurooppalaisissa yhteishankinnoissa, ja aikasarjamallien käytön hyötyjä myös muille yhteishankintayksiköille kuin Hanselille tulisi tutkia lisää.

Avainsanat julkiset hankinnat, keskitetyt hankinnat, aikasarja-analytiikka, ARIMA-malli, kysynnän ennustaminen, puitejärjestely, budjetointi

Acknowledgements

First, I would like to thank Hansel, and especially Kirsi Koivusaari, for giving me the opportunity to be involved in this interesting project. I have learned a lot while writing this thesis, and it has been a pleasure to notice how valuable knowledge and tools I have been able to offer Hansel. Everyone I have worked with during this project has been helpful and eager to support me. In addition, special thanks to Juho Pursimo and Markus Paasiniemi for their invaluable help with the models and comments about the whole thesis. It has been a pleasure to work with you.

Secondly, great thanks to my supervisors Katariina Kemppainen and Katri Kauppi who pushed me to do my best and question my choices. I would have not been able to cover all the aspects of this complex topic without their support.

Lastly, I would like to acknowledge the support I have received from my family and friends. Their numerous comments and ideas have greatly helped me improve the final version of this thesis. It has not always been an easy process, but you never stopped believing I could do it.

Table of Contents

List of Tables

List of Figures

1 Introduction

Making decisions is never easy, especially when they concern the future of a company. Strategic decisions relate to the means of achieving long-term aims and interests (Oxford Dictionary, 2018) and they are "among the main means through which management choice is actually effected" (Papadakis *et al.*, 1998, p. 116). Therefore, strategic decisions often affect the revenue and profits of a company. To illustrate strategic, profit related decision problems, consider the following situation.

"The board of a company is considering a profit warning for the next year. Forecasting the profit for near future, such as in a three-month-radius, is rather accurate, but the farther the forecast reaches, the more it deviates from the actual profit. Therefore, a company can never give completely accurate prediction of the profits. Nevertheless, the board must establish plausible forecasts to reassure the owners and other stakeholders. Moreover, the prediction intervals must be realistic enough for the future profits to fall into them. Hence, after setting the profit objectives based on the forecasts, the board needs to make several strategic decisions to stay within the prediction intervals and meet the long-term profit objective. These decisions include, among others, adjusting the sales prices, optimising the cost structure, and reviewing potential market niches."

This thesis focuses on a such profit decision problem in the field of public procurement. Moreover, the problem is addressed from the perspective of budgeting. A budgeting process is a part of the core functions of a company as without knowledge of incomes and costs it is impossible to plan the future of a company and maintain profitable business. Hence, budgeting supports greatly profit related decision making since being a business process it is "a complete, dynamically coordinated set of activities or logically related tasks that must be performed to deliver value to customers or fulfil other strategic goals" (Trkman, 2010, p. 125). Concluding, the scope of this thesis is to forecast the monetary volume of centralised governmental procurement with time series models and to research whether such models help the case company enhance its budgeting process and better achieve its zero profit objective.

1.1 Motivation

Quantitative techniques, such as analytics, can be used to improve most business functions (Davenport, 2006). Unfortunately, all the data organisations have is still too rarely used to inform business decisions and create positive results (Davenport *et al.*, 2001). This is also the case for the case company, which provides the basis for the motivation of this thesis.

Public procurement has a significant societal role since "purchasing of goods and services […] supports all functions of government" (Thai and Grimm, 2000, p. 232) and it "can play a substantial role in delivering government objectives and […] be a lever for socio-economic development" (Knight *et al.*, 2012, p. 16). Also, public procurement is a notable part of world economy as it accounts for 15 to 20 percent of global GDP and 29 percent of total government expenditure across the OECD countries (Flynn, 2018; European Commission, 2018).

In addition to its economical and governmental importance increased government purchase volumes and the complexity of governmental procurement have made the focus on public procurement higher than ever before (Thai and Grimm, 2000; Knight *et al.*, 2012). Still, there is notably less research conducted on public procurement compared to private procurement, both across and within nations (Karjalainen, 2009; Knight *et al.*, 2012). Moreover, the research that has been done is usually either limited to a specific aspect or in a documentary form (Knight *et al.*, 2012). Also, the previous research focuses mainly on public procurement in general (see e.g. Thai and Grimm, 2000; Erridge and McIlroy, 2002; Lloyd and McCue, 2004; Knight *et al.*, 2012), whether to centralise or decentralise it (see e.g. McCue and Pitzer, 2000; Karjalainen, 2009; Albano and Sparro, 2010) or on some specific phenomenon which can be influenced with public procurement, such as corruption or sustainability (see e.g. Søreide, 2002; Brammer and Walker, 2011).

The shortcoming of using analytical tools and methods in public procurement research is rather surprising, considering how beneficial they can be. Problems faced in both public and private sector necessitate rigorous and representative models as they are complex and often difficult to formulate (Matopoulos *et al.*, 2016). Furthermore, making purchasing decisions at both strategic and operational level can be assisted with empirically driven models (Matopoulos *et al.*, 2016). Fortunately, an increasing number of purchasing management issues are being addressed with analytical tools (Acharya *et al.*, 2009).

The case company – a Finnish central purchasing body Hansel Ltd (later Hansel) – operates in the context of public procurement. In Finland, parts of the procurement of the central government are centralised to a state-owned private company. The functions, objectives, and status of Hansel are defined in the Finnish Act on Hansel (1096/2008, HE 147/2018), and the purpose of the company is to enhance the usage of public funds and to increase the productivity of governmental procurement while being a non-profit organisation. The objectives are met by contracting framework agreements on the joint purchases according to the EU Procurement Directive (Directive 2014/24/EU) and by offering procurement expert and consultancy services related to governmental procurement.

Hansel's revenue is comprised of service fees from the framework agreements and other services. In this thesis only the framework agreements are considered because of their remarkable importance to Hansel and Finnish governmental procurement and the vast data available on their monetary volume. As Hansel is a non-profit organisation, estimating the magnitude of the service fees needs to be accurate enough to be able to cover all costs without making any profit. The service fees are on average 1 percent of the framework agreement volumes (Hansel, 2018a), ranging from 0.5 to 1.5 percent. Since different framework agreements have different service fees, affecting the profit by changing the level of the service fees is strategic, and accurate estimates of the different framework agreement volumes are needed to change the service fees with the most impact. Additional difficulty is the legal aspect of framework agreements; the service fees need to be changed according to the price alteration clauses. Therefore, decisions on service fee changes are not effective immediately.

Hence, the board of Hansel is expecting better forecasts of the framework agreement volumes to better achieve zero profit with adequate service fee allocation instead of artificially increasing costs. In addition, the personnel involved in the budgeting of the framework agreement volumes feels the current process is time consuming, labour-intensive, and inefficient, and that the budget is based more on subjective predictions than facts. Also, the accuracy of the process is unsatisfactory. Even though the upper-level budget target is usually met quite precisely, there is significant scattering between both the framework agreement and the customer dimensions of the budget. This complicates Hansel's objective of zero profit since the service fees from the framework agreements are agreement specific.

The current budgeting process is a combination of top-down and bottom-up budgeting. The total budget target comes from the top management, but the lower-level category and account managers can influence the budgets concerning their framework agreements. During the process the same persons adjust the budget multiple times, and overlapping work clearly exists. Regardless of the inefficiencies, the top management has wanted to uphold the process because of its participatory features. However, in reality the category and the account managers consider the process as unmotivating since the budget levels set by them can be changed quite radically to meet the upper-level target and yet, the category and the account managers are responsible for meeting the budgeted levels in the end. Therefore, the possibilities to influence the budget are in fact specious for the category and the account managers.

The problems of the current budgeting process stem from inadequate utilisation of data and tacit knowledge. To be able to make use of data and analytical tools, companies must first have proper data management practices. Hansel has comprehensive data on the framework agreements for over a decade, but it has not been used in the budgeting process as effectively as it could be. At the moment, no analytical model is used to support the process and the know-how of in-house analysts is not utilised. Also, the tacit knowledge of the category and the account managers is overlooked as meeting the upper-level target is prioritised over achieving an accurate budget on framework agreement and customer levels. Therefore, it can be stated that the project management, i.e. "the analysis and permanent improvement of interdisciplinary tasks" (Rutte, 1990, p. 325), of Hansel's budgeting process has not necessarily been the most effective. Improving the utilisation of both data and tacit knowledge will benefit Hansel as there is positive correlation between business success and project management (Trkman, 2010).

In conclusion, the empirical motivation for this thesis comes from the case company and its need for a more accurate and less laborious budgeting process to better allocate its service fees. This supports the theoretical motivation which, in turn, stems from two shortages in previous research. First, there is an overall need for more research on public procurement, to which this thesis answers by addressing the central government's centralised procurement in one nation. Secondly, this thesis approaches governmental procurement through analytical modelling, which is an approach that has been called for in the procurement literature in general (Matopoulos *et al.*, 2016).

1.2 Objectives and research questions

The aim of this thesis is to research whether time series models will enhance a budgeting process in the context of public procurement. The theoretical aim is to identify if governmental procurement volumes display such behaviour which can be captured and accurately forecasted with time series models. For ARIMA models to produce effective forecasts the estimated data needs to have certain characteristics. Hence, this thesis studies whether data on governmental procurement has these characteristics and how accurately can ARIMA models forecast governmental procurement with the said characteristics and behaviour.

The empirical aim, in turn, is to create usable prediction models for Hansel's budgeting process and thus improve achieving the zero profit objective through more accurate framework agreement service fees. With the models Hansel would be able to forecast the monetary volumes of its framework agreements more accurately and hence, set adequate service fees to reach its non-profit objective. In addition, the models would reduce the complexity and the inefficiencies of the current budgeting process.

Based on the objectives this thesis aims to answer the following research questions:

1. How well can ARIMA time series models capture the behaviour of the framework agreement trade volumes of the central government's centralised procurement?

2. Will these models improve the accuracy of forecasting future framework agreement trade volumes compared to Hansel's current budgeting process?

The focus of this thesis is on the central government's centralised procurement, which is emphasised in the first question. This is an important notion since centralised government procurement may have characteristics which are not necessarily generalisable to all public procurement. The second question complements the first one by considering the usefulness of the models to Hansel and their role in improving the forecasts of the governmental procurement volumes.

1.3 Research approach

This thesis is a case study with a single case company and a quantitative approach. The study can be divided into two parts, a literature review and an empirical data analysis. In the literature review a base for the study and its scope is introduced through previous research, first on public procurement and then on analytics used to support decision making, and more precisely time series models. In addition, a comprehensive description is given of the status of the case company in the Finnish public procurement and of the problem it is facing.

The depiction of the current budgeting process in Chapter 2 is based on interviews of six people at different levels of the company and with varying roles in the process. The interviewees include the controllers, one of the two analysts, the Heads of Units in Category Management, and the Chief Category Officer. Through the interviews the problems of the process are depicted from multiple viewpoints and various wishes for improvement are remarked, including the usage of forecasting models in improving the budgeting accuracy. Hence, also a qualitative perspective is present in the study. The interview questions, the interviewees, and the dates of the interviews are listed in Appendix A.

The empirical data analysis, in turn, strives for creating efficient and accurate prediction models for forecasting the central government's centralised procurement. It is based on the data Hansel has on the monetary volumes of its framework agreements for a 10-year time period. Hansel collects the data monthly through a web portal where the suppliers are expected to report their framework agreement trade. In the data Hansel refers to similar framework agreements with an upper-level concept "framework agreement subtotal" and further "product category". For simplicity and because of the limitations of this thesis, not all framework agreements were included in the study and the selected ones were handled on the subtotal level. The reasoning behind these decisions and how the selection was conducted are explained in more depth in Chapter 4.

Based on the theory of autoregressive integrated moving average (ARIMA) time series models, the data is analysed and modified if necessary and the prediction models are formed. The fitness and reliability of the models are confirmed with statistical tests accustomed when using ARIMA models. When analysing and modifying the data, insights from the category and the account managers is used to explain and exclude possible anomalies.

The empirical part of this thesis is accomplished with the R coding language. The data is exported from Hansel's reporting system and transformed into an RDS data format. All the analyses and modifications are done to this exported data file enabling the original data to stay intact in the reporting system. R is also used for creating and testing the models as well as for the visualisations of the data and the results.

1.4 Structure of the thesis

This thesis is structured as follows. The literature review is divided into two sections covered in Chapters 2 and 3. Chapter 2 focuses on public procurement and the role of the case company in it. First, the general characteristics of public procurement are introduced. Then the focus is shifted on public procurement under the legislation of the European Union and further in Finland. Thirdly, the case company and its role in Finnish governmental procurement is introduced and an overall understanding of the current budgeting process of the case company and its shortcomings is given.

Chapter 3, in turn, concentrates on data analytics used to support decision making and further time series analytics. The general features of time series models are concisely discussed as well as why ARIMA models were chosen to be used in this study, how they are estimated and used for forecasting, and how to evaluate their fitness. Finally, examples of the usage of time series models in previous literature are introduced.

Chapter 4 presents the data and the methods used in model creation. The selected framework agreements and the reasoning behind their selection are gone through as well as the ways and reasons for modifying the concerned data. In addition, the creation of the models and testing their fitness are presented.

The results are introduced in Chapter 5. First, the chosen models and their forecasting accuracy compared to the budget are gone through. Then the results are addressed from the perspective of the research questions and reflected on a more general level. Lastly, Chapter 6 concludes the thesis by summarising the research and its key results and introducing the managerial implications. Also, the limitations of this thesis are recognised and suggestions for future research given.

2 Case company in public procurement context

As Hansel is a state-owned company aiming to reduce public expenditure through increasing the productivity of governmental procurement (Hansel, 2018a), it acts in the context of public procurement. Therefore, to understand Hansel and the challenges it faces in its budgeting process, one needs to comprehend what public procurement is and how it is implemented in Finland. In this chapter, a concise overview on public procurement in general and further in the European Union and in Finland is given. Then the focus is shifted on Hansel, its role in the Finnish public procurement field, and its current budgeting process.

2.1 Public procurement

Procurement as such refers to professional-like acquiring of goods and services, including all stages of the process from determining the need to contract completion and closeout (Lloyd and McCue, 2004). Further, public procurement refers to a systematic and somewhat strategic procurement made by public authorities. Over time, different terms have been used of the buying function of government organisations, including “purchasing”, “contracting”, and “acquisition” (Lloyd and McCue, 2004). Public procurement is the term used by academic and legal journals (Lloyd and McCue, 2004) and it can be seen as an upper-level definition for all the different terms since “it refers to the acquisition of goods and services by governments or public sector organisations through a public contract” (Witjes and Lozano, 2016, p. 38). This is also the definition this thesis is based on.

Public procurement has a long history, dating back to around 2500 B.C. (see e.g. Thai and Grimm, 2000; Thai, 2001), and a central role in public service delivery (Flynn and Davis, 2014). In the past, public procurement has been seen as an instrument involved in securing national economic and social policies (McCrudden, 2004), and the focus has been on individual contracts or transactions (Lloyd and McCue, 2004). Also, previous research has mostly engaged with practitioners and addressed their interests without relation to theoretical perspectives (Flynn and Davis, 2014). During the past decades the focus has shifted, however, to more strategic public procurement (Lloyd and McCue, 2004) and further, to a new area of sustainable public procurement – a combination of green and social procurement – to address both social and environmental issues (McCrudden, 2004).

All governmental units need goods and services to carry out their operations and procurement officials to efficiently and effectively manage procurement (Thai and Grimm, 2000). Hence, public procurement supports governments to achieve economic, social, and other objectives such as economic growth, environmental sustainability, and social inclusion (Thai, 2001; Knight *et al.*, 2012; Flynn and Davis, 2014) in an equitable, transparent, and economical way (Thai and Grimm, 2000). Moreover, efficient procurement is essential for national development (Kumar *et al.*, 2017).

The goals of public procurement can be further divided into two groups: procurement and non-procurement goals (Thai, 2001, p. 27). The procurement goals include overall costs, quality, timeliness, financial and technical risks, minimising business, maximising competition, and maintaining integrity. The non-procurement goals, in turn, include economic goals, social goals, environment protection, and international relations goals.

In addition to the goals, there are certain expectations set for public procurement by both society and legislation. Knight *et al.* (2012, p. 17) have identified these additional demands for public procurement and categorised them based on their sources and effects. These additional demands are (1) external demands set for the righteousness of procurement, (2) internal demands concerning the public organisation, (3) demands originating from the context, (4) demands on the process, and (5) multiple roles for the public organisation itself. These demands are further depicted in Table 1.

Table 1. Demands set for public procurement by society and legislation (Knight et al., 2012)

External demands	• transparency • integrity • accountability • exemplary behaviour
Internal demands	• many goals at the same time • political goals • many stakeholders
Demands originating from the context	• budget driven • open budget • mutually dependent budget situations • cultural setting
Demands on the process	• strict limits from legal rules and organisational procedures • long-term relationships with suppliers • cooperation with other public entities
Multiple roles for the public organisation	• large buyers • reciprocity with the suppliers • determiner of the rules and regulations

Furthermore, public procurement has some interests it needs to balance with. According to Knight *et al.* (2012, p. 19) these interests are primary, secondary, process, and competition interests. Primary and process interests are the same in both public and private procurement whereas secondary and competition interests differ across the sectors. Primary interests refer to the buyer's interest to purchase the materials or services needed and to the supplier's interest of continuity of their organisation and of profit making. Process interest, in turn, is the interest of both the buyer and the supplier to minimise the transaction costs of the process. Secondary interests refer to deeper interests of the buyer and supplier than the primary interests. For the supplier, these can be for example increasing market share or gaining experience, while for a public procurer secondary interests are related to the objectives and responsibilities of government in general. Lastly, competition interest is the interest of public buyers and suppliers to have competition and to make sure the competition is fair.

To meet with all the objectives set for public procurement, interdisciplinary skills and knowledge are required (Thai, 2001) as well as an efficient and functional procurement system. Even though it is impossible to create a one-size-fits-all public procurement system since the organisational structures vary with the size of government units (Thai, 2001), there are some guidelines for a sound one. Traditionally a centralised purchasing authority has been seen as a way to assure economy, efficiency, and integrity in public procurement, but this notion has been challenged since the turn of the century and replaced by a structure involving more sharing and delegation (McCue and Pitzer, 2000; Thai and Grimm, 2000).

In general, procurement systems can be divided into centralised, mid-range, and decentralised systems. According to McCue and Pitzer (2000), in a centralised system the lines of authority and functional responsibility are clearly defined, and a central purchasing department is in charge of decision making and control whereas the roles of line departments are limited to requesting goods and services. A mid-range system is a combination of centralised and decentralised system in which a given purchasing department is liable for policy making and overseeing responsibility as well as facilitating the use of the purchasing authority granted to the line departments, typically on a dollar limit base. In a decentralised system, the enabling legislation and purchasing policies are the only central authority, and line departments are both responsible and accountable for the success or failure of their purchases. In practice, procurement systems are different combinations of centralized and decentralized processes (McCue and Pitzer, 2000), and thus, an idealised system does not exist, as already mentioned. Nonetheless, it is essential that throughout the system well-defined authorities and responsibilities exist for each level of management (Thai, 2001) to enable achieving the many objectives of public procurement.

Considering all the objectives, goals, tasks, and responsibilities public procurement has, it is only natural that there are several laws and regulations concerning it. In fact, Lloyd and McCue (2004, p. 3) have stated that "public procurement is one of the most highly legislated and regulated fields of government". One reason for this is that public procurement is one of the government activities most vulnerable to corruption (Thai, 2001). Hence, governments need to ensure healthy competition by avoiding contact with suppliers prior to the publication of tenders (Witjes and Lozano, 2016). Another aspect is the balancing of governments between participating in the market as purchasers and at the same time regulating it through their purchasing power (McCrudden, 2004). All in all, to prevent unethical and illegal business practices it is necessary to have a public procurement system with clearly stated goals and policies which are implemented either by non-legal means, such as internal administration, or by formal rules and regulations (Thai and Grimm, 2000; Thai, 2001). However, given that there is no one uniform procurement system, there are also fundamental differences between countries in regulatory and legislative arrangements (Knight *et al.*, 2012). Therefore, public procurement is next viewed specifically from the viewpoint of the European Union legislation and then more precisely in Finland.

2.1.1 Public procurement in the European Union

Public procurement has long suffered from definitional ambiguities and lack of uniformity across state and local governments (Lloyd and McCue, 2004). This combined with the previously mentioned differences in procurement systems and regulations makes comparing public procurement across countries difficult. However, the European Union has managed to establish a somewhat coherent public procurement environment for the European Economic Area (EEA) with common legislation. The importance of unified regulation to effective public procurement is emphasised by Lloyd and McCue (2004), who have stated that even though the United States is one nation, it is less of a common market than Europe because of the lack of uniform public procurement rules.

Public procurement in the European Union accounts for around 14 percent of GDP (European Commission, 2018). It is regulated by the Directive of the European Parliament and of the Council on public procurement. The directive was first introduced in 2004 and revised in 2014 with Directive 2014/24/EU to increase the efficiency of public spending with particular interest on the participation of small and medium-sized enterprises in public procurement and to enable procurers to make better use of public procurement in supporting common societal goals. In the directive, rules for procurement procedures of contracting authorities with respect to public contracts and design contests are established. Nevertheless, the directive does not affect all public purchases, only the ones with a value net of value-added tax (VAT) estimated to be equal or greater than the thresholds specified in the directive, ranging from EUR 134 000 to EUR 5 186 000. Furthermore, these thresholds are verified every two years to correspond to the thresholds established in the World Trade Organisation Agreement on Government Procurement. (Directive 2014/24/EU)

The directive deals with multiple aspects of public procurement from definitions and transparency to preparation and award of contracts, to name a few. Nonetheless, the most important concept of the directive, in the context of this thesis, is a framework agreement. It is defined as "an agreement between one or more contracting authorities and one or more economic operators, the purpose of which is to establish the terms governing contracts to be awarded during a given period, in particular with regard to price and, where appropriate, the quantity envisaged" (Directive 2014/24/EU, Chapter II, Article 33).

Framework agreements differ from public contracts by establishing the terms which govern the contracts that are awarded during the agreement period (Andrecka, 2016). Since their introduction in 2004, framework agreements have gained popularity and importance on the EU public tender market and have become a key element of public procurement in several countries, including Finland, Denmark, Sweden, the UK, and France (Andrecka, 2016). However, there are both conveniences and problems with framework agreements, introduced by Andrecka (2016). On one hand, with framework agreements every public purchaser does not have to go through lengthy EU procurement processes, which increases procurement efficiency. Also, commercial benefits are gained by aggregating buying power and a multitude of suppliers helps manage risk. On the other hand, framework agreements have the potential to be bid-rigging, limit the access of small and medium-sized enterprises (SMEs) to public tenders, and close the market for competition. Moreover, framework agreements lock authorities out from new technologies occurring on the market, since the supplier offering the best technology when the agreement is concluded may not be a technology leader at the moment of the call-off. This is problematic as making public procurement more strategic and spurring innovation is one of the goals of the directive (Andrecka, 2016).

Along with the legally-binding directives, the European Commission issues policy guidance about public procurement. After Flynn (2018, p. 3) this includes recommendations and advice on "how to best achieve value of money (VfM), facilitate SMEs in contract competitions, source innovative product and service solutions, and promote environmental and social objectives through ethical purchasing". The purpose of policy guidance is to achieve public procurement which plays a strategic role in public service delivery and contributes to an economically prosperous, financially stable, socially inclusive, and environmentally sustainable Europe (Flynn, 2018).

The European Commission has commendable goals with its directives and policy guidance, but have they been actualised in practice? The Commission measures procurement performance across the EEA members with six VfM indicators (European Commission, 2015):

1. one bidder, i.e. does the buyer have a choice between suppliers
2. no calls for bids, i.e. the measure of openness and transparency in advertising and award of contracts
3. aggregation, i.e. exploitation of economies of scale with more than one contracting authority procurement procedures
4. award criteria, i.e. price-based criteria or a mix of price and quality factors
5. decision speed
6. reporting quality

Based on the recent data from Tenders Electronic Daily (TED), the official listing site for public contracts in the EEA, each country is classed with above average, average, or below average performance (Flynn, 2018). The procurement performance of each EEA country across the six indicators are listed in Table 2. Unfortunately, no country has satisfactory performance across all the indicators, but Finland comes close with only "one bidder" being average. Also, all the Scandinavian countries, the Benelux, the Anglo countries, and Malta have above average performance. Nonetheless, there are slightly more countries with below average performance than above average, and the performance of the recent EU member countries lags behind that of the founding or long-standing member countries (Flynn, 2018).

Table 2. Procurement performance of the EEA countries (Flynn, 2018)

	One bidder	No calls for bids	Aggregation	Award criteria	Decision speed	Reporting quality
Austria	≈	✓	X	✓	✓	X
Belgium	≈	✓	✓	✓	✓	X
Bulgaria	X	≈	X	✓	X	✓
Croatia	X	X	X	X	✓	✓
Cyprus	X	X	✓	X	✓	✓
Czech Republic	≈	X	X	X	✓	✓
Denmark	✓	✓	✓	✓	✓	X
Estonia	≈	X	X	X	✓	✓
Finland	≈	✓	✓	✓	✓	✓
France	≈	✓	X	✓	✓	X
Germany	≈	≈	X	✓	✓	X
Greece	≈	✓	X	X	X	✓
Hungary	X	X	✓	✓	✓	✓
Iceland	✓	≈	✓	✓	✓	X
Ireland	✓	✓	X	✓	X	X
Italy	X	≈	✓	✓	X	X
Latvia	X	X	✓	X	✓	✓
Lithuania	≈	≈	X	X	✓	✓
Luxembourg	✓	✓	X	✓	✓	X
Malta	≈	✓	✓	X	✓	✓
Netherlands	✓	✓	X	✓	✓	X
Norway	≈	✓	✓	✓	✓	X
Poland	X	✓	X	X	✓	✓
Portugal	≈	✓	X	✓	✓	X
Romania	X	X	X	X	✓	✓
Slovakia	X	X	X	X	X	✓
Slovenia	X	X	✓	✓	✓	✓
Spain	≈	X	X	✓	✓	X
Sweden	≈	✓	✓	✓	✓	X
UK	✓	✓	✓	✓	✓	X

✓ = Satisfactory performance
≈ = Average performance
X = Unsatisfactory performance

Provably there are significant performance gaps between the EEA countries, and the desired level of performance has not yet been reached (Flynn, 2018). This is surprising since all the countries are subject to the same regulatory and policy regime. Hence, the differences must derive from the national level. Different levels of implementation of the directive reflect on the national legislations and thus the procurement performances. Additionally, the directive and policy guidance have several objectives which can be conflicting (Halonen, 2016). Contradictory emphasis on the objectives in national regulation may accordingly result in differing procurement practices. Therefore, the relatively strict public procurement legislation of Finland may explain why it ranks the best in procurement performance among the EEA countries. Next, the Finnish way to execute public procurement is introduced.

2.1.2 Public procurement in Finland

In Finland, the annual public procurement has been estimated to equate approximately EUR 30 billion, which is almost 25 percent of the GDP (Confederation of Finnish Industries, 2018). Further, the central government procurement amounts to EUR 4-4,5 billion, excluding infrastructure and defence procurement (Ministry of Finance, 2018b). Yet, the definition of public sector is ambiguous which affects also the definition and computation of the total volume for public procurement. According to the Ministry of Finance (2018a), in Finland the public sector includes the central government, the municipalities and joint municipal authorities, the Provincial Government of Åland, the statutory pension insurance companies and institutions, and other social security funds. Because of the importance of public procurement to Finland’s national economy and the complexity of the public sector compared to the private sector, precise regulation is justifiable.

Public procurement is regulated in Finland by (Ministry of Finance, 2018b)

1. The Act on Public Contracts and Concessions (1397/2016)
2. The Act on Hansel (1096/2008, HE 147/2018)
3. Government Decree on Centralised Procurement of the State (765/2006)
4. The Decision of Ministry of Finance on Centralised Procurement of the State (766/2006)

The Act on Public Contracts and Concessions executes the Directive 2014/24/EU as well as Directives 89/665/EEC, 2007/66/EC, and 2014/23/EU, and composes a Finnish vision of how public procurement should be carried out. The contracting authorities obliged to obey the Act on Public Contracts and Concessions are defined in the aforementioned act, and include the central government, the municipalities and joint municipal authorities, the Evangelical Lutheran and the Eastern Orthodox church, unincorporated government enterprises, and public institutes (Valtion hankintakäsikirja, 2017).

In addition to the legislation, the Ministry of Finance has published *Handbook on Government Procurement* (2017) in co-operation with Hansel. The handbook is based on the Directive 2014/24/EU and the Act on Public Contracts and Concessions, and it aims to "support the implementation and organisation of procurements by central government procurement units as well as planning and tendering of procurements, the conclusion of procurement contracts and their follow-up" (Valtion hankintakäsikirja, 2017, p. 5). Even though the handbook is especially targeted to the procurement units of the central government, it can be utilised where applicable by other public procurement units as well.

The Act on Public Contracts and Concessions has three objectives (Hansel, 2018b):

1. To enhance the usage of public resources.
2. To promote the quality of procurement.
3. To secure the opportunities of companies to offer goods and services to the government and the municipalities.

The act obligates the government and the municipalities to use public tenders in procurement and defines in detail practices and rules related to procurement, such as different tendering means (Hansel, 2018b). In addition, the national thresholds, ranging from EUR 60 000 to EUR 500 000, are specified in the act (Valtion hankintakäsikirja, 2017, p. 75). Based on them, purchases are divided into three categories:

1. Small purchases falling below the national thresholds.
2. National purchases exceeding the national thresholds but falling below the EU thresholds.
3. EU purchases exceeding the EU thresholds.

Only the national and EU purchases fall under the Act on Public Contracts and Concessions (Valtion hankintakäsikirja, 2017). Even though the small purchases are beyond the regulation of the act, suggestions of good practices in procuring them are given, and also the general principles of EU – transparency, objectivity, non-discrimination, and relativity – should be considered during the procurement process (Valtion hankintakäsikirja, 2017).

The Finnish public procurement legislation has been criticised for overemphasising transparency (Halonen, 2016). Ensuring open and undistorted competition as well as developing effective competition are some of the fundamental purposes of the EU public procurement rules (Halonen, 2016), and they are achieved with the right amount of transparency. In fact, transparency is an important anti-corruption tool and needed for effectiveness. Nevertheless, unnecessary transparency increases the risk of competition distortions and facilitates the formation of bid-rigging cartels (Halonen, 2016). As there are no general, explicit rules for transparency in the EU directives (Halonen, 2016), the level of transparency is determined in the national legislations. Therefore, Finland has adopted interpretation of full transparency in accordance with the Act on the Openness of Government Activities (Halonen, 2016). In the act, it is stated that all official documents are in the public domain in order to promote openness and good practice of information management in government and to provide private individuals and corporations the opportunity to monitor the exercise of public authority and the use of public resources (621/1999).

Nonetheless, according to Flynn (2018), Finland has outstanding public procurement performance compared to the rest of Europe when assessed with the six VfM indicators. Hence, the strong emphasis on transparency has not threatened public procurement practices. It may, however, affect negatively both the national and the EU markets as the fear of needing to disclose strategic information and trade secrets decreases the desire of companies to take part in public tenders (Halonen, 2016). Therefore, the balance between fair competition and transparency of the government must be maintained also in the future.

2.2 Hansel

Through legislation Finland has centralised public procurement to central purchasing bodies (CPBs). In principle, Finland has two CPBs, one for the central government and one for the municipalities and joint municipal authorities. In practice though, there are several contracting authorities procuring for the municipalities. At the time of writing, the CPB for the central government is Hansel and for the municipalities KL-Kuntahankinnat Ltd. However, a merger of these two CPBs is expected to take place during 2019. To enable the merger the Act on Hansel was revised in December 2018 so that also the municipalities and joint municipal authorities can be Hansel's customers. Nonetheless, this thesis focuses on the current situation of Hansel with only the central government as its clientele.

As mentioned, Hansel operates under the guidance of the Finnish government, more precisely under the Ministry of Finance (Hansel, 2018a). This sets stricter objectives and procedures for Hansel compared to common private companies. Perhaps the clearest distinction is the non-profit objective. Since around 87 percent of Hansel's revenue comes from the framework agreement service fees, their significance for the non-profit objective is remarkable. The service fees are adjusted according to budgeted framework agreement volumes which gives the budgeting process a key role in achieving zero profit. Even though the service fees can be changed either within the agreement period in accordance with the price alteration clauses of the framework agreements or during a tendering process of a new agreement, the effects are not imminent, and the role of accurately budgeted volumes is enhanced. Hence, the focus of this thesis is on improving the accuracy of the current budgeting process.

Next, Hansel's role in Finnish governmental procurement is introduced to get a general idea of the operational environment of the company. Then, an overall depiction of the current budgeting process and its deficiencies as well as suggestions for improvement are given based on the interviews carried out with six of Hansel's employees.

2.2.1 Governmental procurement in Finland

Public and governmental are often used as synonyms in public procurement literature. However, there is a distinct difference between these two forms of procurement. Public procurement covers procurement done by all governing agencies, such as municipalities and joint municipal authorities as well as public enterprises and universities. Governmental procurement, in turn, only consists of the procurement of the central government and agencies and institutions under its direct authority. In Finland, the scope of governmental procurement is defined in the Act on Hansel by determining which public contracting authorities can use Hansel's services, and therefore be included in the central government procurement. Consequently, the central government procurement in Finland covers the agencies and institutions under the budgetary economy of the central government, the Finnish Parliament and units under its authority, contracting authorities owned by, funded by, or under the mandate of the central government as well as unincorporated government enterprises (1096/2008).

As a CPB, Hansel provides centralised purchasing activities for the Finnish central government. The reason for centralising the central government's procurement is to enhance the usage of public resources by increasing the productivity of governmental procurement (Hansel, 2018a; HE 147/2018). This is in line with the statement of McCue and Pitzer (2000, p. 404) that a central procurement system develops purchasing expertise and through that "increases efficiency and economy and insures the integrity of the purchasing system". In fact, with the usage of framework agreements the Finnish government has been able to reduce overlapping tendering work as well as achieve remarkable price benefits by pooling large volumes of government purchases since the founding of Hansel (HE 147/2018). With the framework agreement volumes totalling to EUR 800 million in 2017, these price benefits equal approximately to EUR 235 million per year, or a 20.5 percent savings compared to decentralised purchasing (Karjalainen *et al.*, 2008; Hansel, 2018a).

Also, many other European countries have similar CPBs (HE 147/2018), listed in Table 3. Although the core function of all European CPBs is to tender joint purchases for their clients, many of them have also other functions related to procurement and the clienteles vary between countries (HE 147/2018). The most significant difference with Hansel and other European CPBs is that Hansel's clientele covers only agencies and institutions under the budgetary economy of the central government as well as contracting authorities funded by or under the mandate of the central government. Other European CPBs for their part serve also the contracting authorities of regional governments and municipalities in addition to the central government (HE 147/2018).

Table 3. European Central Purchasing Bodies

Country	Central Purchasing Body	Established	Employees
Austria	BBG	2001	113
Belgium	CMS	2002	5
Bulgaria	CFCU	2010	21
Croatia	State Office for Central Public Procurement	2010	43
Denmark	SKI	1994	80
Finland	Hansel	2003	94
France	UGAP	1985	1200
Germany	Beschaffungsamt	1951	220
Hungary	KEF	1997	-
Iceland	Ríkiskaup	1949	25
Ireland	Office of Government Procurement	2013	195
Italy	Consip	1997	240
Norway	SI (GPC)	2016	10
Portugal	ESPAP	2012	14
Slovenia	Public Procurement Directorate	2012	33
Spain	DGRCC – Minhap	2013	84
Sweden	Kommentus	2011	48
Sweden	SIC	2011	50
UK	CCS	2014	700

The duties of Hansel are specified in the Act on Hansel. It acquires goods and services to the contracting authorities of the central government by awarding public procurement contracts and concluding framework agreements (1096/2008). In addition, Hansel provides expertise and development services related to procurement procedures to the contracting authorities, such as customer specific tendering, as well as collecting and analysing procurement information as a part of governmental procurement digitalisation programme (HE 147/2018).

Concluding and managing framework agreements is the most remarkable function of Hansel. Centralising procurement with framework agreements has grown popularity which can be seen from the increased volume of procurement channelled through Hansel’s framework agreements (Lempinen, 2013). In fact, in 2017 the volume of framework agreement purchases were over 6 times higher than 10 years before as the framework agreement trade was worth EUR 800 million compared to EUR 140 million in 2007 (Hansel, 2018a; HE 147/2018). In Figure 1 is depicted the framework agreement volumes as well as the number of subtotals, product and service categories, and suppliers related to the framework agreements for the past 12 years. At the moment, Hansel has around 70 framework agreement subtotals within 17 product and service categories, and almost 400 suppliers.

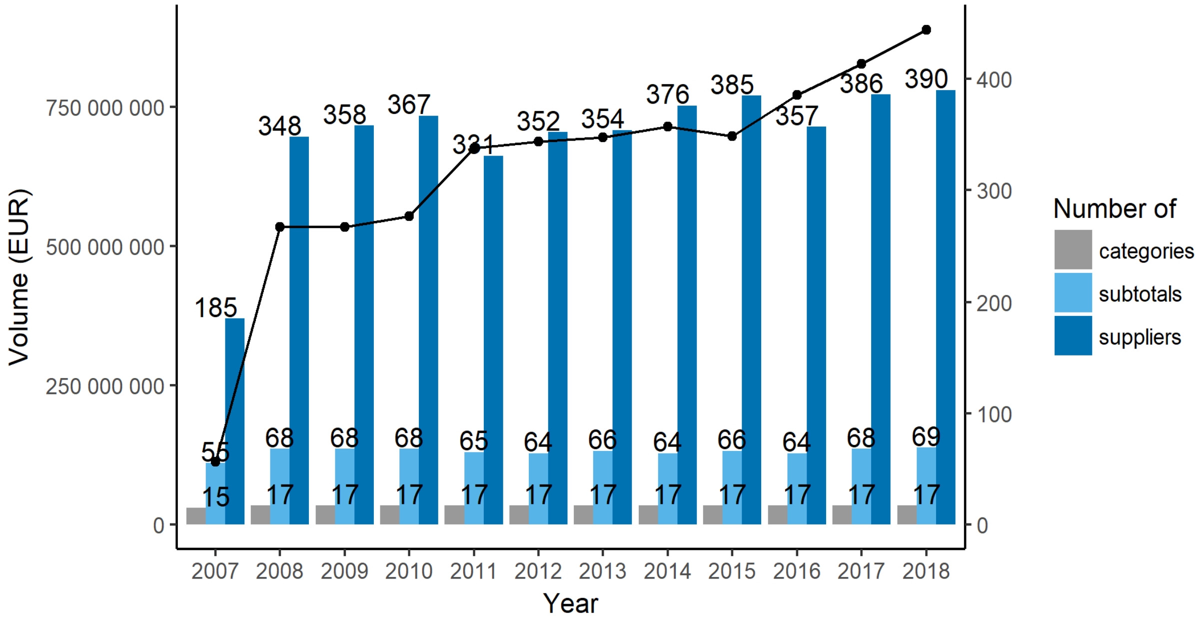


Figure 1. Development of Hansel's framework agreement trade over the past 12 years

Andrecka (2016) points out the benefits of framework agreements concluded by CPBs. Such framework agreements have larger capacity and include usually larger quantity and variety of products than the ones established by a single contracting authority. Furthermore, there are more users involved, the value of the framework agreements is high, and there is usually a greater level of professionalism in carrying out the procurement (Andrecka, 2016). Despite the benefits, the compliance rates of Hansel's framework agreements are only 20-80 percent, even though the obligation of the central government to implement procurement through framework agreements is decreed in the Act on Government Budget (423/1988) (Karjalainen, 2009; HE 147/2018). The low compliance rates are mainly caused by agency problems, i.e. information asymmetry and goal incongruence, between the operational buyers and the purchasing units (Karjalainen, 2009). Hence, there is further potential to increase the framework agreement purchases and the consequent cost savings to the Finnish government.

2.2.2 Budgeting process in Hansel

The budgeting process of the framework agreement volumes in Hansel is a combination of top-down and bottom-up budgeting. In top-down budgeting lower-level management develop their budgets based on the global framework plans and guidelines set by the top management (Kramer and Hartmann, 2014) whereas in bottom-up budgeting most budgeting decisions are initiated by lower-level officials and the decision making is decentralised

(Hendrick, 1989). There are pros and cons in both approaches, and the choice of budgeting style needs to be done on a case-by-case basis. In general, top-down budgeting enhances the economic exchange relationship managers have with the organisation (Kramer and Hartmann, 2014). Moreover, top-down budgeting results in lower agency costs and budgetary slack while setting expected compliance and targets for the lower-level management (Heinle *et al.*, 2013; Kramer and Hartmann, 2014). In turn, bottom-up budgeting supports the social exchange relationship of managers with the organisation by encouraging them to influence their day-to-day activities and performance targets (Heinle *et al.*, 2013; Kramer and Hartmann, 2014). As a down side, managers may benefit from misreporting in the absence of expected budget target levels (Heinle *et al.*, 2013).

In Hansel's budgeting process, the budget target comes from the top level, and the lower-level managers can influence the framework agreement and customer specific budgets, as depicted in Figure 1. The process starts with the CEO, the CFO, and the Chief Category Officer (CCO) composing an upper-level budget target for the next year based on the forecast of the on-going year. This forecast is a combination of the realised volume until July and the budget of the rest of the year. After this the category and the account managers set budget levels for each framework agreement according to previous volumes and the insight they have into the procurement behaviour of the customers. If these budget levels lack behind the upper-level target, the CCO modifies the framework agreement budgets together with the Heads of Units in Category Management. Then the controllers allocate the framework agreement budgets to the customers according to the distribution of trade in the on-going year. Once again, the budget goes to the account managers who can change the customer-specific budgets if needed before the CCO and the Chief Customer Officer approve the framework agreement budgets. The controllers comprise a budget for the service fees based on the framework agreement budgets and the current service fees, and the CFO and the Chief Accountant prepare a budget for costs separately. Final adjustments are made to the budgets by the CEO before the board approves the overall budget.

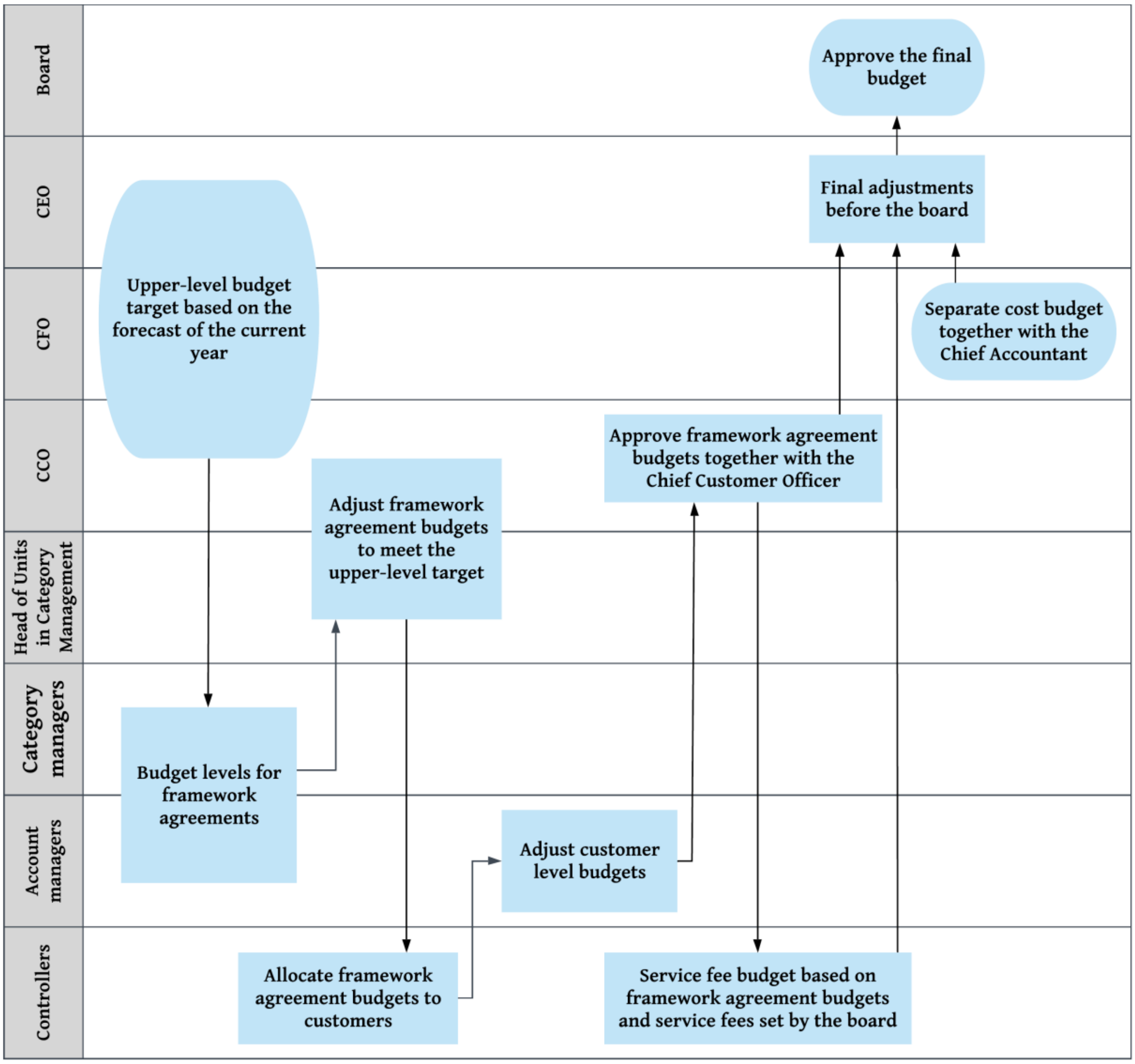


Figure 2. Current budgeting process in Hansel

The current budgeting process has several deficiencies resulting from the following facts:

1. The budgeting process is inefficient with time-consuming and repetitive manual labour.
2. The budgeting process is based on subjective predictions and assertions instead of facts.
3. The possibilities of the lower management to influence the budget are specious.
4. The persons involved in the budgeting process have differing definitions for a budget.

As the process is labour-intensive with manual and time-consuming steps, it is rather inefficient. Also, there are repetitive stages as the CCO as well as the category and the account managers adjust the framework agreement budgets multiple times. If the resulting budget was highly accurate, all the effort and time put into the process could be justifiable, but that is not the case. In Table 4 are depicted the percentage differences between the actual and the budgeted volumes on the product and service category levels. The upper-level budget target is usually met quite precisely as the total has differed from the budget on average 4.8 percent over the past 5 years. However, there is significant scattering between the categories, which is caused by inaccurate budgets for both the framework agreements and the customers. The good performance of the upper-level budget is at least partly caused by the aggregation of the framework agreement and customer level inaccuracies; the opposite differences between the budgets and the actual volumes cancel each other out.

Table 4. Differences between the actual and the budgeted volumes in percentages over the past 5 years

Category	**2017**	**2016**	**2015**	**2014**	**2013**
Vehicle Services	-1,6 %	40,9 %	8,5 %	8,5 %	36,6 %
Professional Services	-32,5 %	44,0 %	25,3 %	13,7 %	-2,4 %
Energy	-10,3 %	-11,0 %	-16,5 %	-17,7 %	-11,4 %
Human Resources and Health Services	3,8 %	1,9 %	-0,8 %	4,3 %	3,9 %
IT equipment	22,1 %	-10,8 %	-29,6 %	17,2 %	16,4 %
Data Centre Services and Equipment	37,4 %	42,8 %	25,9 %	11,0 %	-26,4 %
Consulting Services	28,8 %	33,5 %	23,4 %	15,2 %	-1,6 %
Transport and Logistics Services	-16,1 %	-23,9 %	2,3 %	15,4 %	13,1 %
Consumer Products and Consumables	6,7 %	4,4 %	-2,1 %	22,0 %	6,0 %
Accommodation and Conference Services	9,0 %	-1,2 %	-13,9 %	6,9 %	2,8 %
Travel Services	5,9 %	2,0 %	-1,6 %	2,7 %	9,2 %
Software	17,5 %	6,7 %	-6,1 %	-3,7 %	8,2 %
Financial Services	23,5 %	0,3 %	-23,6 %	7,9 %	36,3 %
Telecommunications	18,8 %	-2,5 %	-10,1 %	7,7 %	3,6 %
Office Services	4,2 %	-3,3 %	-8,2 %	-4,6 %	0,7 %
Facilities Management Services	-0,7 %	19,9 %	-1,7 %	7,0 %	12,7 %
Security Technology and Services	-48,8 %	31,0 %	72,7 %	117,8 %	-32,7 %
Total	**5,9 %**	**5,6 %**	**-5,9 %**	**2,8 %**	**3,7 %**

The budgeting inaccuracies on the framework agreement and the customer levels are a direct consequence of adjusting the budgets set by the category and the account managers to correspond to the upper-level target. The account managers know their customers and their procurement needs best, but this tacit knowledge is lost during the process. In addition, the account managers rarely adjust the budgets allocated to the customers since if they increase or decrease a certain budget, the corresponding amount has to be added to or subtracted from somewhere else. Therefore, the bottom-up aspect of the current budgeting process is artificial since the framework agreement budgets set by the category and the account managers are adjusted arbitrarily by the upper management and the possibilities to influence the final budget are scarce.

Another deficiency of the current process is confusion between different operators about the definition of a budget. From the category and the account managers point of view the budget is more like a target or a sales goal as the upper management dictates future volumes which the category and the account managers try to reach during the coming year. This results in a motivation problem when the performance of the category and the account managers is evaluated by how well they meet the budget but their insight into the possible future volumes is overlooked. The CEO and the CCO, in turn, sees the budget as more of a prediction of the coming year's volumes which is then approved by the board and expected to actualise.

All these inefficiencies and contradictions in the budgeting process affect making informed and effective decisions about the service fee levels. The service fees are framework agreement specific, and decisions about increasing or decreasing them are done based on the budgeted framework agreement volumes. Since the budget is more of an assertion of the top management than an actual prediction with robust fact base, the differences between the budgeted and the actualised framework agreement volumes make it almost impossible to make accurate updates to the service fees and through them achieve the non-profit objective.

The problems of the current process have been noted in Hansel and several expectations for improvement have been expressed. First, a clear distinction between a target and a prediction should be made. Additionally, Hansel's budget for the framework agreement volumes is in essence a sales budget. Hence, to achieve an accurate prediction, the customer units should comprise the budget which would then be reflected to higher levels through the category managers. Secondly, the utilisation of technology and the level of automation during the budgeting process is low. The framework agreement and customer level budgets are done in Excel sheets which are complicated to use and contain too little background information about the allocated numbers. Therefore, a system-based approach would be desired to increase the transparency of the process and help simulate different scenarios for the future volumes. Also, automating the process to a high degree would free the time of Hansel's experts to other, more prominent tasks, such as providing tendering consulting to the contracting authorities of the state. Thirdly, the potential calculation should be used to support the budgeting to be aware of the maximum framework agreement volumes for the next year. The potential calculation considers on a customer level the framework agreement volumes, personnel, purchase invoices, and other customer specific information for the previous year to calculate the achieved volume if the customer used only framework agreements for all its purchases.

At the moment, no model is used to support the budgeting process even though Hansel has comprehensive data on framework agreement volumes for over a decade. If time series models were utilised during budgeting, they would be most beneficial when used by the account managers. Nonetheless, since utilising such models requires statistical and technical knowledge, a very low-effort user interface would be needed. Therefore, the models would most likely be managed by the analysts at the beginning of the budgeting process. Then the estimates of the framework agreement volumes produced with the models would be adjusted by the category and the account managers according to their tacit knowledge. This way the budgeted volumes would be better allocated to correspond to the actual volumes, and they would be based on facts rather than subjective predictions.

3 Time series analytics in decision making

Using data and quantitative analysis are among the most powerful tools for improving decision making (Davenport, 2009), such as in the case of Hansel’s service fees. Indeed, a growing number of companies is basing their decisions on data analytics as the improvements in technology have broadened the approaches to decision making. According to Davenport (2009), decisions based on analytics are more likely correct and more justifiable because of the rigour of scientific methods. Nevertheless, Davenport (2009) also remarks that it may be difficult and time-consuming to gather enough usable data and to keep it up to date. In addition, almost all quantitative models are based on historical data, which may not always be the best indicator for the present and the future.

Furthermore, not all analytics answer to the same questions. Analytics can be divided into five levels based on the objectives and the level of automation utilised, as depicted in Figure 3 (Dykes, 2017). Descriptive analytics uses automation the least, and it is the most common type of analytics used by companies (Davenport, 2006). Using data to describe what happened and diagnosing why it happened is rather easy and, in many situations, informative for the company. But to improve decision making, analytics should be able to give suggestions of how a company should react to possible future outcomes. This can be achieved by increasing the level of proficiency in analytics and utilising predictive and prescriptive analytics (Davenport, 2006).

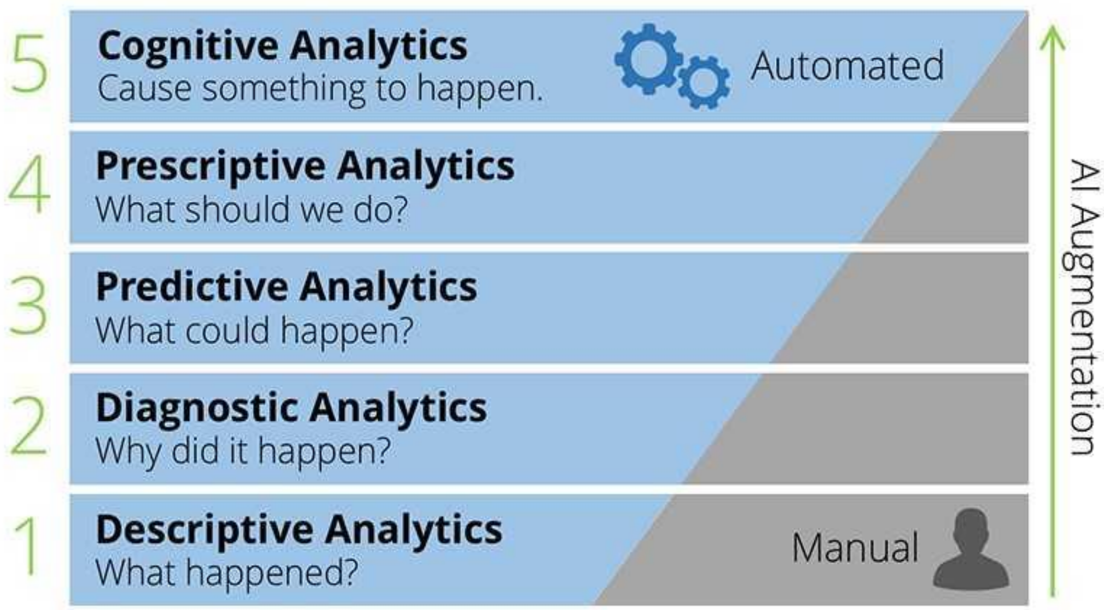


Figure 3. Different levels of analytics (Dykes, 2017)

This thesis focuses on predictive analytics as the objective is to forecast the future volume of framework agreements and make decisions based on the forecasts. Further, the scope is on time series analytics since the monthly data on the framework agreement volumes has observations at an equally spaced interval of time, and therefore, it equals to a time series (Box *et al.*, 2008).

Forecasting is commonly the goal of time series analytics (Tsay, 2000), and it provides a solid basis for decisions made, for example, in production planning, finance and risk management, and industrial processes (Box *et al.*, 2008; Montgomery *et al.*, 2015). There are three basic approaches to generating forecasts within time series analytics: regression models, smoothing models, and time series models (Montgomery *et al.*, 2015). Regression models predict the variable of interest through its relationship with one or more related predictor variable (Montgomery *et al.*, 2015), and of all statistical methods, regression is one of the most widely used (Ruppert and Matteson, 2015). However, in this case regression models are not suitable for forecasting the framework agreement volumes since the volumes are not expected to be affected by other variables than themselves.

Smoothing models use typically a simple function of previous observations to forecast the variable of interest (Montgomery *et al.*, 2015). Unfortunately, they fail to take advantage of the serial dependence of adjacent observations, an essential feature of time series, in the most effective way (Box *et al.*, 2008; Montgomery *et al.*, 2015). Time series models, in turn, are formal models which are based on the statistical properties of historical data and which incorporate the dependent structure of observations (Montgomery *et al.*, 2015). Therefore, time series models depict the true behaviour of the variable of interest better than smoothing models, and that is why they were chosen for forecasting the framework agreement volumes.

In general, time series models used for forecasting are stochastic models, or probability models, as they depict a sequence of random variables, also called a stochastic process (Box *et al.*, 2008; Ruppert and Matteson, 2015). A wide class of stochastic processes is provided by a range of models called autoregressive integrated moving average (ARIMA) models, which represent many of the time series met in practice (Box *et al.*, 2008). In addition, there are multiple other time series models such as generalised autoregressive conditional heteroscedasticity (GARCH) models, vector autoregressive (VAR) models, and vector error correction models (VECM) (see e.g. Lütkepohl, 2005; Box *et al.*, 2008; Ruppert and Matteson, 2015) which are beyond the scope of this thesis.

ARIMA models are used for forecasting in this thesis because in addition to efficiently capturing the dependent structure of time series, they are a class of "time series models with only a finite, preferably small, number of parameters" (Ruppert and Matteson, 2015, p. 314). This makes estimating the models and forecasting the framework agreement volumes fast, straightforward, and rather accurate. Next, the qualities required from the data and possible data transformations needed to be able to use ARIMA models are gone through. Also, selecting and estimating such models are introduced. Then forecasting with ARIMA models and ensuring the forecast accuracy are covered. In addition, examples of time series models used in previous literature are introduced at the end of this chapter.

3.1 ARIMA models

Autoregressive integrated moving average (ARIMA) models are a combination of autoregressive (AR) and moving average (MA) models with possible integration to remove nonstationarity. AR models explain the temporal dependence of a variable of interest with a finite, linear aggregate of past observations, i.e. lags, and a white noise error term (Box *et al.*, 2008; Ruppert and Matteson, 2015). MA models, in turn, depict the variable of interest as a weighted average of all past values of the white noise error term (Ruppert and Matteson, 2015). The definition of white noise is presented in Appendix B.

The objective of creating any time series models is to keep the lag structure simple, i.e. achieve a parsimonious model, but at the same time depict the process as accurately as possible. AR models are the simplest models for depicting stochastic processes, but there is a potential need for multiple lags to capture all the dependency of the variable when using only AR processes (Ruppert and Matteson, 2015). To avoid this and to achieve a parsimonious model, an MA component is added to the AR process resulting in an ARMA model (Ruppert and Matteson, 2015). If the time series in question is nonstationary, an order of integration is needed to create a stationary process, and the model in question becomes an ARIMA model. Integration and when it is used are further covered in Chapter 3.1.3. The definition and meaning of stationarity is, in turn, introduced in Chapter 3.1.1. Concluding, ARIMA models can be used for depicting and forecasting various time series, both stationary and nonstationary, effectively and accurately (Ruppert and Matteson, 2015).

The simplest AR model, the first-order autoregressive model AR(1), uses only the first lag to explain the dependence of the variable of interest as shown in equation (1) where y_t is the variable of interest, δ is a parameter corresponding to a constant, ϕ captures the temporal dependence, and ε_t is the white noise error term (see e.g. Montgomery *et al.*, 2015). A general representation of an AR process is the *p*th-order autoregressive model, AR(p), given in equation (2), in which the last *p* values of the process explain the variable of interest y_t (see e.g. Ruppert and Matteson, 2015).

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t \tag{1}$$

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{2}$$

Equation (2) can also be expressed as the first-order system or in a matrix form (see Appendix B), which is helpful with further computations as *p* increases.

As a comparison to the AR(1), the first-order moving average model MA(1) explains the dependence of the variable of interest as given in equation (3) where δ corresponds to a constant, θ captures the temporal dependence similarly to ϕ in equation (1), and ε is the white noise error term (see e.g. Montgomery *et al.*, 2015). The general *q*th-order representation MA(q) depicted in equation (4) uses the current and the *q* previous lags of the error term to explain the variable of interest. The MA(q) in equation (4) can also be expressed as the first-order system or in a matrix form (see Appendix B) to help further computations, similar to AR(p).

$$y_t = \delta + \varepsilon_t - \theta \varepsilon_{t-1} \tag{3}$$

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{4}$$

Therefore, an ARIMA(p,d,q) model corresponds to equation (5) or more simply to equation (6) which utilises the matrix forms of AR(p) and MA(q) in equations (B.3) and (B.4), respectively. The d in the model definition corresponds to the order of integration needed to ensure a stationary process.

$$y_t = \delta + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{5}$$

$$Y_t = \Delta + \Phi Y_{t-1} + E_t - \Theta E_{t-1} \tag{6}$$

Next the properties of ARIMA models along with model estimation and selection is covered briefly. Then, forecasting with ARIMA models and evaluating the forecast accuracy is introduced. The mathematical representation of each property, model estimation, and forecasting is kept scanty since the theory behind them is not in the focus of this thesis but rather their practical implications.

3.1.1 Stationarity and invertibility

One of the most useful methods to achieve the most accurate representation of a time series with minimum number of lags, i.e. a parsimonious model, is "to assume some form of distributional invariance over time, or stationarity" (Ruppert and Matteson, 2015, p. 307). Stationary models assume that the unconditional moments of the process do not change over time (Box *et al.*, 2008). Mathematically this means that "the probability distribution of a sequence of n observations does not depend on their time origin" (Ruppert and Matteson, 2015, p. 308). Therefore, strict stationarity requires that all aspects of the behaviour of a process are time-invariant. However, it is a very restrictive assumption, and hence, weak stationarity, in which the mean, variance, and autocovariance of a process are independent of time, is usually a sufficient requirement (Ruppert and Matteson, 2015). The mathematical representation of weak stationary is depicted in Appendix B.

In AR models, stationarity is linked to the parameters capturing the temporal dependence. An AR(1) model is stationary when $|\phi| < 1$ in equation (1) (Box *et al.*, 2008). Likewise, the absolute value of each ϕ_i where $i = 1, \dots, p$ in equation (2) needs to be smaller than 1 for an AR(p) to be stationary. Then the unconditional moments – mean, variance, and autocovariance – are defined, finite, and time-invariant (see Appendix B). Because of the more complex lag structure, obtaining the unconditional moments for an AR(p) is most convenient with the matrix form in equation (B.3).

As MA(q) processes use only the lags of the white noise error term, they are always stationary (see Appendix B). However, they have an invertibility condition to put higher weight on the most recent error term values (Hyndman and Athanasopoulos, 2018), and thus, to make forecasting possible. Invertibility is independent of stationarity and it is also applicable to nonstationary models (Box *et al.*, 2008). Nevertheless, like stationarity, invertibility is linked to the parameters capturing the temporal dependence in MA(q) models.

An MA(1) is invertible when $|\theta| < 1$ in equation (3), and an MA(q) when all $|\theta_i|$ where $i = 1, \ldots, q$ in equation (4) are smaller than 1 (see e.g. Box *et al.*, 2008).

Furthermore, an ARIMA(p,d,q) is stationary and invertible when the eigenvalues of both Φ and Θ in equation (6) lie inside the unit circle (see Appendix B). Having a stationary AR(p) process has few important implications. First, the forecast converges geometrically quickly to the mean the farther we go into the future (Ruppert and Matteson, 2015). In other words, the long-run forecast corresponds to the unconditional expectation in equation (B.8) for a stationary process. Secondly, the dynamic responses, i.e. the effect of a current shock on the current and future values of a process, have only a transitory effect. This means that their effect decays exponentially as the time distance from the moment of the shock grows (Montgomery *et al.*, 2015). Computing dynamic responses is not in the interests of this thesis, and therefore, not covered here.

3.1.2 Autocorrelation and partial autocorrelation

Autocorrelation and partial autocorrelation are useful properties for detecting the stationarity and invertibility of ARIMA processes, respectively. As autocovariance measures the covariance between two values y_t and y_{t-i} i lags apart, autocorrelation depicts the relation of autocovariance to the changes in lag size (Box *et al.*, 2008). In other words, an unconditional autocorrelation measures the temporal dependence of the variable of interest. Partial correlation, in turn, means "the correlation between two variables after being adjusted for a common factor that may be affecting them" (Montgomery *et al.*, 2015, p. 359). Partial autocorrelation simply refers to the partial correlation of the variable of interest with itself.

For a stationary AR(1) the unconditional i-order autocorrelation equals to ϕ^i (Box *et al.*, 2008). As $|\phi| < 1$ when AR(1) is stationary, the unconditional autocorrelation tends rapidly to zero when $i \to \infty$ (see e.g. Ruppert and Matteson, 2015). The properties of AR(1) models are easily generalised to AR(p) models (Ruppert and Matteson, 2015), and hence, the sample autocorrelation function (ACF) can be used to detect whether an AR(p) process is stationary (see e.g. Box *et al.*, 2008; Ruppert and Matteson, 2015). Therefore, it can be concluded that for all stationary AR(p) processes the unconditional autocorrelation tends asymptotically to zero, and these processes are said to have a short memory (Box *et al.*, 2008).

Further, ACF can be used to detect if a stationary process is white noise. As the variables of a white noise process are independent, the unconditional autocorrelations are zero and approximately normally distributed (Montgomery *et al.*, 2015). The Ljung-Box test is a statistical test for checking if a process is white noise. If the Q statistic in equation (B.16) is larger than the $\chi^2(k)$ critical value, the null hypothesis of the sample autocorrelations of order *i* being jointly zero is rejected, and the process is not white noise.

Partial autocorrelation, in turn, is related to the invertibility of MA(q) processes. The i-order partial autocorrelation associated with an invertible MA(1) equals to $-\theta^i$ (Box *et al.*, 2008). Therefore, the partial autocorrelation tends to zero when $i \to \infty$ as $|\theta| < 1$. Like with AR processes, also the properties of MA(1) models can be generalised to MA(q) models. Hence, the partial autocorrelation function (PACF) can be used to identify an invertible process, since it tends exponentially to zero for invertible MA(q) processes (see e.g. Box *et al.*, 2008).

Autocorrelation and partial autocorrelation can also be used to identify the lag structure of MA and AR models, respectively. The ACF of an invertible MA(q) process has a cut-off after q, i.e. it is zero or nearly zero for lags larger than q implying that an MA(q) is a proper fit for the process. Similarly, the PACF of a stationary AR(p) has a cut-off after *p* so that the values of PACF at larger lags are near to zero. This in turn indicates that an AR(p) is a proper fit for the time series in question. (Box *et al.*, 2008; Ruppert and Matteson, 2015)

3.1.3 Trends and seasonality

Often transforming time series data is useful to make it easier to interpret the results, to stabilise the variance of the data, and to achieve stationarity (Montgomery *et al.*, 2015), among other things. Expressing time series in logarithm or in percentage helps interpreting the results, and the log transformation is an optimal transformation for stabilising variance when the standard deviation increases linearly with the mean (Montgomery *et al.*, 2015). In addition to transformations, several types of adjustments are utilised in time series modelling. Two of the most widely used ones, trend and seasonal adjustments (Montgomery *et al.*, 2015), are covered here.

A time series which exhibits a trend is nonstationary (Montgomery *et al.*, 2015) since its unconditional expectation, i.e. the mean, is not constant and time-invariant. Such time series can be divided into a trend component and a cyclical component as in equation (B.13) from which the trend can be removed with a linear regression model (see Appendix B). In this case, however, the trend is assumed to be deterministic and it needs to be estimated. Therefore, using differencing is more advantageous than model fitting, since it does not require estimation and allows the trend component to change through time (Montgomery *et al.*, 2015). In differencing a new time series is obtained by applying a difference operator to the original time series to acquire the differences of the subsequent values of the variable of interest (see Appendix B). Differencing corresponds to the integration part of ARIMA models, and it is used in general to remove nonstationarity. Differencing can be executed as many times needed, but in practice, one or two differences is usually enough to remove nonstationarity (Montgomery *et al.*, 2015).

Another common component causing nonstationarity is seasonality. Strong seasonal variation is often exhibited in economic time series (Ruppert and Matteson, 2015) and it can be removed such as a trend. Even though also regression models can be used to eliminate seasonality (see e.g. Montgomery *et al.*, 2015) seasonal differencing is most often used (Ruppert and Matteson, 2015). Seasonal differencing resembles regular differencing but instead of capturing the difference between the subsequent values of the variable of interest, the seasonal difference operator depicts the difference of each value to the value s lags before (see Appendix B). Many time series have both a seasonal and a trend component, in which case seasonal differencing is first used to remove the seasonality and then the trend is removed by differencing one or more times (Montgomery *et al.*, 2015).

3.1.4 Parameter estimation

There are several methods for parameter estimation, such as moments, least squares, and maximum likelihood (Montgomery *et al.*, 2015). However, most ARIMA models are nonlinear and therefore, they need to be estimated with nonlinear model fitting procedures like conditional least squares and maximum likelihood (Montgomery *et al.*, 2015; Ruppert and Matteson, 2015). The estimation of ARIMA(p,d,q) models has two steps. First, the time series is differenced d times to achieve stationarity. Then, the parameters of the remaining ARMA(p,q) are estimated with conditional least squares or maximum likelihood.

Both methods are based on the log-likelihood function, i.e. the logarithm of the joint density function (see e.g. Ruppert and Matteson, 2015). With maximum likelihood, the parameter estimators are obtained by maximising the log-likelihood function. Conditional least squares, in turn, maximises the logarithm of the conditional density function. Nowadays, parameter estimators can be automatically computed by most statistical software packages (Ruppert and Matteson, 2015) so only the basic idea of both methods is introduced in Appendix B to give a general perception.

3.1.5 Model selection

In many cases there are several plausible models for a time series and additional analysis is needed to find the best one. According to Montgomery *et al.* (2015, p. 61), the basic steps in modelling a time series include (1) plotting the series to detect its basic features such as trends and seasonality, (2) using data transformations and eliminating trends and seasonality if necessary, (3) estimating the plausible models, and (4) validating the performance of the models from the previous step to select the model with the best fit.

Plotting a time series allows preliminary detection of stationarity, trends, and seasonality. If the plotted series exhibits trend or seasonality, they need to be eliminated by differencing before fitting a model. In addition, the stationarity and invertibility of a series can be confirmed by examining its autocorrelation and partial autocorrelation functions. ACF reveals whether a process is stationary and the lag structure of an MA(q) process as it tends asymptotically to zero for stationary processes and has a cut-off after lag q when a process is an MA(q). Similarly, PACF tends asymptotically to zero for invertible processes and has a cut-off after lag p suggesting an AR(p) process (see e.g. Montgomery *et al.*, 2015).

After determining the possible lag structure few different models are estimated with, for example, conditional least squares or maximum likelihood, and their goodness of fit is tested (Box *et al.*, 2008). A widely used test to compare the goodness of fit of two or more models is the likelihood ratio test depicted in equation (B.25). However, the maximised value of log-likelihood used in the test can be increased by adding parameters to the model, which does not necessarily mean better fit (Ruppert and Matteson, 2015). Hence, “to find a parsimonious model one needs a good tradeoff between maximising fit and minimising model complexity” which is achieved with Akaike’s (AIC) and Bayesin (BIC) information criteria (Ruppert and Matteson, 2015, p. 109). Both information criteria, introduced in equations (B.26) and (B.27), are based on the log-likelihood, and selecting the model with the smallest AIC or BIC leads to a parsimonious model with good fit.

Lastly, it is needed to confirm that the residuals of the selected model are white noise. The residuals ε_t of an ARMA(p,q) should not have significant autocorrelation, if the model fits the time series well (Ruppert and Matteson, 2015). Conversely, if the residuals of the selected model have autocorrelation, the selected lag structure fails to capture the temporal dependency of the variable of interest extensively enough. Confirming that the residuals are white noise is done by examining their ACF and by using the Ljung-Box test in equation (B.16) (Ruppert and Matteson, 2015). If there are multiple models with white noise residuals and no clear distinction between their goodness of fit, the most parsimonious model is selected.

3.1.6 Forecasting and forecast accuracy

Forecasting means "predicting future values of a time series using the current information set" containing present and past values of the series (Ruppert and Matteson, 2015, p. 342). There are ready functions for forecasting in most statistical software packages, but to give a general idea, the central concepts of forecasting with ARIMA(p,d,q) models are introduced. These include the *i*-step-ahead forecast, the *i*-step-ahead forecast error, and the variance of the forecast error (see Appendix B). For a stationary process, the forecasts tend to the unconditional expectation in equation (B.8), i.e. the mean, and the variance of the forecast errors tends to the unconditional variance in equation (B.9) as *i* increases (Ruppert and Matteson, 2015). The converging of the variance increases the uncertainty associated with the forecast as time horizon increases.

Specifying the accuracy of the forecasts is necessary to, for example, calculate the risks associated with decisions based on the forecasts (Box *et al.*, 2008). The standard ways to measure forecast accuracy are the mean error (ME), the mean absolute error (MAE), and the mean squared error (MSE) depicted in equations (B.31) - (B.33). However, the ME, MAE, and MSE are scale-dependent, i.e. their values are expressed in terms of the original units, which might not be informative enough in some cases (Montgomery *et al.*, 2015). Therefore, the relative forecast error (RE) in equation (B.34) is used to accomplish comparison across different time series or time periods (Montgomery *et al.*, 2015).

An important notion is to measure the forecast accuracy using data which was not used in the modelling. Generally, the best fit to historical data does not result in the best forecasts of new data (Montgomery *et al.*, 2015). Therefore, using the same data sample for fitting the models and measuring the forecast accuracy causes inaccurate results about the prediction capabilities of the models. Hence, the method of data splitting is used when evaluating the forecast accuracy of time series models (Montgomery *et al.*, 2015). In-sample data is used for modelling and forecasting i-step-ahead forecasts where i corresponds to the size of out-of-sample data set. Then the forecast accuracy is measured by comparing the forecasts and the out-of-sample data to evaluate how well the models can forecast unknown data, i.e. how usable are they for forecasting future values of a time series. The best approach to forecast a time series with ARIMA models is to select the model with the smallest MSE from the out-of-sample forecast (Montgomery *et al.*, 2015).

3.2 Time series models in previous literature

As mentioned, data analytics is increasingly utilised in decision making, and time series analytics is a clear subset of predictive data analytics. To get an understanding of different uses of time series analytics, few examples from the previous literature are introduced. Even though the examples are from various fields, almost all of them are related to some sort of decision making situations which emphasises the role of time series analytics in decision making.

Specific decision making processes are covered by Hahn *et al.* (2009) and Ediger and Akar (2007). They study the use of time series models for load forecasts in electricity sector and for forecasting energy demand in Turkey, respectively. According to Hahn *et al.* (2009, p. 902) the decision process in the electricity sector is “complex with several different levels that have to be taken into consideration” and at the core of the process is finding appropriate approach and model for load forecasts. Therefore, they cover different models and methods traditionally used for forecasting load demands of which time series approaches, such as ARMA, ARIMA, and seasonal ARIMA models, are among the oldest ones. Ediger and Akar (2007, p. 1701), in turn, concentrate on estimating the future primary energy demand in Turkey with ARIMA and seasonal ARIMA models as “forecasting energy demand in emerging markets is one of the most important policy tools used by decision makers all over the world”.

Cheung *et al.* (2003), Chandra and Al-Deek (2009), Tandberg and Qualls (1994), and Dekimpe and Hanssens (2000) address decision making more indirectly in the fields of customer service management, traffic prediction, marketing, and emergency healthcare.

Cheung *et al.* (2003) propose for customer service management (CSM) a multi-perspective knowledge-based system (MPKBS) which incorporates an AR model to predict the adapted value of the quality of service, and through that to monitor the performance of the customer service staff continuously. Chandra and Al-Deek (2009), in turn, improve the traditional ARIMA models used for freeway traffic prediction by utilising VAR model to demonstrate the effect of upstream and downstream locations on traffic at specific locations. Applications of MA and ARIMA models in the field of healthcare are introduced by Tandberg and Qualls (1994) as they use the said time series models to predict emergency department volume, length of stay, and acuity. Lastly, Dekimpe and Hanssens (2000) give an overview on past research which utilises various time series methods, including ARMA and VAR models, in marketing.

To address the topics of this thesis more closely, examples from the fields of budgeting and procurement are considered further. Reddick (2002) compares three budgetary decision making models which use time series analytics on real disaggregated national government budget outputs. A simple random walk, i.e. a nonstationary AR(1), represents garbage can budgeting, whereas the budgetary incrementalism model implies that governmental public spending follows a stationary AR(1). Lastly, in rational choice budgeting government surplus evolves from a strict random walk process for the first difference of the surplus. This model can be further represented as an MA(1). Shahandashti and Ashuri (2013) address the problem of variations in the construction cost index (CCI) for cost estimation and budgeting of capital projects. Previously seasonal ARIMA models and Holt-Winters exponential smoothing (ES) have been proposed for forecasting CCI, but Shahandashti and Ashuri show that several VECM are more accurate than the previous models. Ilbeigi *et al.* (2017, p. 1), in turn, create four time series models – ES, Holt-Winter ES, ARIMA, and seasonal ARIMA – to “take into account the short-term variation of asphalt cement price in forecasting its future values” in order to improve the budgeting of transportation projects.

Wang *et al.* (2018), Acharya *et al.* (2009), and Kumar *et al.* (2017), for their part, consider time series models in procurement literature. Wang *et al.* (2018, p. 212) investigate “how to optimally plan bunker procurement using the swap contract to hedge the procurement risk” and create a calibrated multivariate GARCH for describing the movements of the swap contract price and the spot market price of the bunker. Acharya *et al.* (2009), in turn, present a VAR model to establish the empirical relationship between product prices at different regional markets and fuel costs. They conclude that their VAR modelling approach can also be used “in virtually any situation where products can be sourced from multiple locations” (Acharya *et al.*, 2009, p. 224). Finally, to examine co-integration and the subsequent short-run and long-run effects of adopting a new public procurement policy in Fiji on the aid inflows from bilateral donors Kumar *et al.* (2017) use the autoregressive distributed lag (ARDL) procedure.

Even though time series analytics and models are widely used methods in previous literature, only few examples were found to consider public procurement or procurement in general. Of the three studies addressing procurement with time series models only Kumar *et al.* (2017) focused specifically on public procurement. Moreover, Reddick (2002) considered national government budget outputs with time series analysis which has close relationship with the topic of this thesis. Nevertheless, this overview of previous literature emphasises the lack of utilising time series models in both private and public procurement research.

4 Data and methods

This chapter describes the empirical data analysis as well as the data and methods used in the creation of the forecast models. First, a forecasting process, according to which the model creation was executed, is introduced. Secondly, the used data and necessary data transformations are presented. Finally, different models are fitted to the data, the final models are chosen for the forecasting process, and their forecasting accuracy is evaluated. Both the handling of the data, the model fitting and forecasting, and the evaluation of the forecast accuracy were done in R with the ready-made ARIMA forecasting documentation and principles by Hyndman and Athanasopoulos (2018).

4.1 Forecasting process

The empirical part follows the forecasting process defined by Montgomery *et al.* (2015, p. 13). In general, the basic elements of forecasting processes are similar since certain steps are required to generate effective forecasts. The process by Montgomery *et al.* was chosen to be used as it is straightforward and comprehensive and consists of 7 distinct steps, described in Figure 4. In (1) problem definition, an understanding of the usage of the forecast as well as the expectations of the user are developed. The step includes determining, among other things, the desired form of the forecast, the forecast horizon and interval, and the required level of forecast accuracy. (2) Data collection means obtaining the relevant historical information for the variable(s) of interest. Emphasising the relevance of the information is the key; often not all historical data is useful for a current problem and it is necessary to deal with missing values or other data-related problems. (3) Data analysis suggests initial types of models to explore through visual inspection of recognisable patterns, such as trends and seasonality, and numerical summaries of the data, including sample mean, standard deviation, and autocorrelation.

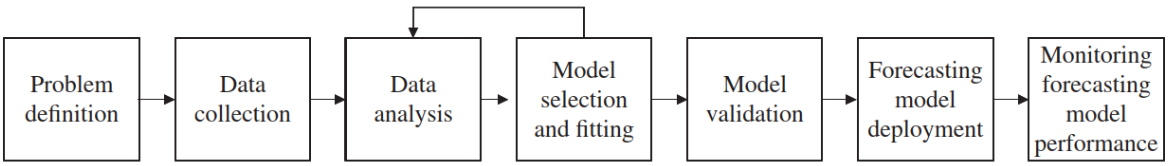


Figure 4. Forecasting process (Montgomery et al., 2015)

(4) Model selection and fitting includes choosing one or more models and fitting them to the data by estimating the unknown model parameters. The goodness of fit to the historical

data of the models is also evaluated in this step. (5) Model validation, in turn, evaluates the ability of the models to forecast new data. This is usually done with data splitting, i.e. using a part of the data for model fitting and another part for forecasting the intact values and then evaluating the generated forecast errors. In (6) forecasting model deployment, the model and the resulting forecasts are put to use. If the user is someone not accustomed to using forecast models, it is important to ensure generating forecasts becomes routine and that data sources and other required information will continue to be available. (7) Monitoring forecasting model performance should be "an ongoing activity after the model has been deployed to ensure that it is still performing satisfactorily" (Montgomery *et al.*, 2015, p. 16). Monitoring forecast errors is essential since conditions change over time, and models may deteriorate in performance even though they have performed well in the past.

The main focus of this study is on steps 1 to 5 since the study does not extend far enough into the future to thoroughly monitor the model performance. Nevertheless, some emphasis is put on model deployment and monitoring model performance through the managerial implications for Hansel. The problem definition has been already covered in Chapter 2 along with the case context introduction. Therefore, the steps from Data collection to Model validation are covered here next.

4.2 Data

The used data is the data Hansel has on the framework agreement trade for about a 10-year period. The data depicts the monthly volumes of the trade made through each framework agreement, reported by the suppliers through a web portal. As the data was not collected by the author specifically for this study but rather by the controllers for Hansel's own purposes, it is classified as secondary data (see Hox and Boeije, 2005). This also alters the Data collection step slightly as the needed data was reconstructed from the existing data to correspond to the requirements and limitations of this thesis.

For confidentiality reasons suppliers, customers, and different framework agreements were not considered in the modelling. The data was grouped into framework agreement subtotals (later subtotals), which combine the volume of framework agreements with similar product and service categories. The models were fitted to these subtotals, and the total volume of the grouped framework agreements were forecasted. This was done so that a broader portion of the framework agreements could be covered on a more general level than if individual agreements were modelled. In addition to simplicity and keeping the number of models manageable, this eliminated the problem of framework agreements lasting usually only for four years (see Directive 2014/24/EU), which is quite a short time period for historical data. Also, forecasting accurately a contract that has just begun is practically impossible, not to mention having no point in forecasting a contract that is about to end. By using subtotals, the distortions caused by ending and starting framework agreements overlapping each other could be cancelled out.

The forecasted subtotals were selected based on the framework agreements that had trade in 2017. Out of the 66 subtotals (see Table C.1) 25 were included in the modelling. Each of them had volume over EUR 10 million, which enabled to cover around 85 percent of the total volume of the framework agreement trade in 2017 with around a third of the subtotals. Moreover, this way the most important subtotals by volume were modelled and the achieved effect on the budgeting process was maximised.

However, some additional modifications to the selected subtotals were made based on the insights and expectations of the CCO and the CFO. Electricity was not included even though it had the second biggest volume in 2017 because of its dependence on the market price and the belief of the category manager that a forecasting model would not considerably benefit predicting the future volumes of Electricity. Human Resources Services, in turn, was included as a partial subtotal as two of its four framework agreements were considered remarkably more important compared to the other two agreements. Further, Facility Services for Premises was divided into Facility Cleaning Services, Facility Security Services, and Restaurant Services since Hansel has great growth expectations for each of them and therefore, wants separate models for forecasting them. Because of this division, Facility Security Services and Restaurant Services did not meet the total volume over EUR 10 million criterion on their own. The final 26 subtotals chosen to be modelled and their total volumes in 2017 are listed in Table 5.

Table 5. Framework agreement subtotals chosen for modelling and their total volumes in 2017

Framework agreement subtotal	Total volume (EUR)
Occupational Health Care Services	61 195 914
Computers	53 658 032
IT Consulting	51 938 885
Data Centre and Capacity Services	40 571 646
Leasing Services	40 054 770
Cars	37 740 538
Microsoft	34 961 045
Scheduled Flights	31 635 638
Facility Cleaning Services	31 350 273
Office Furniture	29 306 001
Fuel	28 417 845
Domestic Accommodation and Conference Services	27 849 414
Teleoperator Services	20 418 796
Management Consulting	16 539 046
Office and IT Supplies	14 354 621
Vehicle Leasing Services	13 058 484
Groceries and Non-food Products	12 863 702
Rail Transport Services	12 101 912
Cell Phones	10 991 517
IT Network Equipment	10 679 539
Printing Services	10 215 322
Security Technology and Services	10 131 988
Human Resources Management Services	10 030 479
Projection Techniques	10 002 158
Restaurant Services	8 597 177
Facility Security Services	7 949 740

4.3 Modelling and model evaluation

The forecasting models were fitted based on the theory of ARIMA models introduced in Chapter 3 and the automatic properties of R. First, each subtotal was plotted to detect trends and seasonality, as stated in Data analysis step. The plotted subtotals are shown in Figure C.1. According to the plotted series and their ACF and PACF, few models were fitted manually with the R function *Arima* (see Hyndman, 2018) and their forecast accuracies were evaluated. To get an unbiased assessment of the forecast accuracies, the method of data splitting was used, and the data was divided into in-sample and out-of-sample data sets. The models fitted to the in-sample data set were used to forecast the same time period ahead that the out-of-sample data set covered. These forecasted values and the out-of-sample data set were then compared to calculate the typical forecast accuracy measurements, such as the mean forecast error and the relative error. It is recommended to have at least 20 to 25 observations in the out-of-sample data set to get adequate results (Montgomery *et al.*, 2015).

Hence, in this case, the out-of-sample data set covered the monthly observations from January 2017 to October 2018, thus containing 22 observations.

To minimise the effect of human error also automatic model fitting was utilised. The R function *auto.arima* fits automatically the best ARIMA model to a univariate time series given to it as a parameter according to information criteria (see Hyndman, 2018). The function also considers trends and seasonality which makes it suitable to be used in this context. Data splitting was used also when fitting models with *auto.arima* to get comparable results with the manually fitted models.

Calculating the forecasts and the forecast accuracy measurements for both manually and automatically fitted models was done automatically with R functions *forecast* and *accuracy*, respectively (see Hyndman, 2018). Final distinctions between the manually and automatically fitted models were made based on their accuracy to forecast the actual subtotal volumes. In cases when the different forecast accuracy measurements produced contradictory results, the selection was made based on the root mean squared error (RMSE), as selecting the model which results in the smallest MSE is in general the best approach (Montgomery *et al.*, 2015).

Finally, the ability of the chosen models to improve the budgeting accuracy compared to the current budget was evaluated. First, the forecast errors of the forecasted and the budgeted volumes were compared to see if the models produced smaller forecast errors than the current budgeting process. In other words, the actual subtotal volumes from January 2017 to October 2018 were compared to both the forecasted and the budgeted volumes of the same time period, and the same forecast accuracy measurements were calculated for both pairs of values. As numerical accuracy does not tell the whole truth, the actual, the forecasted, and the budgeted volumes were then plotted into one graph to determine whether the models captured the behaviour of the actual volumes more accurately than the budgets. Finally, conclusions about the ability of the models to improve the accuracy of the current budgeting process were done based on the differences in both the forecast errors and the graphs.

An example of the modelling code and its outputs is depicted in Appendix D. First, the in-sample and the out-of-sample data sets are created, and the budgeted volumes fetched. Next, the in-sample data is plotted and different ARIMA models are fitted first automatically and then manually based on the needed differencing and the ACFs and PACFs. Then the models which produce white noise residuals are used to generate forecasted volumes. Lastly, the accuracies of the forecasts and the budget are evaluated, and the forecasted volumes of the chosen model are compared to the actual and the budgeted volumes graphically.

Comparing the forecast performance of the models to the budget was time-consuming and laborious with multiple steps to cover. There were 26 subtotals to model, and to each of them several models, about ten on average, were fitted to find the most accurate one. Because of the multiple phases of the modelling the framework agreement volumes were decided to be modelled on the subtotal level; to get as much of the total volume of the framework agreement trade covered as possible with manageable workload. The technical execution of model fitting and forecast accuracy evaluation is not addressed further since the focus of this thesis is on the results and whether the chosen models performed better at forecasting future framework agreement volumes than the current budgeting process. The chosen models and their forecast performance are covered more profoundly in Results.

5 Results

The results of this thesis, i.e. the chosen models and their ability to increase the budgeting accuracy compared to the current budgeting process, are introduced in this chapter. First, the models chosen based on the data and methods introduced in Chapter 4 are gone through. Second, the numerical forecast accuracy of the chosen models is compared to the accuracy of the current budgeting process. Third, the ability of the models to improve the budgeting accuracy in contrast to the current budgeting process is evaluated through graphical inspection. Finally, the results are summarised and their practical implications and effects on future research are discussed.

5.1 ARIMA models chosen for forecasting

The models chosen for each of the 26 subtotals are listed in Table 6 along with the needed data transformations. In the column “Model” the chosen model is reported for each subtotal in standard ARIMA(p,d,q)(P,D,Q) form, in which the brackets indicate the non-seasonal and the seasonal part of the model, respectively. The p and P refer to the AR terms, the q and Q to the MA terms, and the d and D to the order of integration, i.e. differencing. Therefore, for example the ARIMA(0,1,1)(0,0,3) model of Computers has no AR terms (p = 0 and P = 0) but one non-seasonal and three seasonal MA terms (q = 1 and Q = 3). In addition, only the first difference has been taken (d = 1 and D = 0). In practice, the non-seasonal AR and MA terms are simply the p and q previous lags of y_t and ε_t in equation (5), whereas the seasonal AR and MA terms refer to the P and Q previous lags of the same variables starting from the sth lag. As stated in Appendix B, the value of s depends on the type of the time series, and in this case $s = 12$ as the framework agreement trade is reported monthly.

Table 6. ARIMA models chosen for forecasting the framework agreement subtotals

Framework agreement subtotal	Differencing	Model
Occupational Health Care Services	1st and seasonal	ARIMA(0,1,3)(2,1,0)
Computers	1st	ARIMA(0,1,1)(0,0,3)
IT Consulting	1st and seasonal	ARIMA(0,1,1)(0,1,1)
Data Centre and Capacity Services	seasonal	ARIMA(1,0,0)(0,1,1)
Leasing Services	1st	ARIMA(2,1,1)
Cars	seasonal	ARIMA(0,0,1)(0,1,1)
Microsoft	seasonal	ARIMA(3,0,0)(1,1,0) with drift
Scheduled Flights	seasonal	ARIMA(0,0,1)(0,1,2)
Facility Cleaning Services	1st	ARIMA(0,1,1)(1,0,1) with drift
Office Furniture	1st	ARIMA(1,1,1)(2,0,0)
Fuel	seasonal	ARIMA(1,0,3)(1,1,1)
Domestic Accommodation	1st and seasonal	ARIMA(2,1,0)(0,1,1)
Teleoperator Services	1st	ARIMA(1,1,4)
Management Consulting	1st and seasonal	ARIMA(3,1,1)(1,1,1)
Office and IT Supplies	none	ARIMA(1,0,1)(1,0,3) with mean
Vehicle Leasing Services	1st	ARIMA(2,1,3)
Groceries and Non-food Products	seasonal	ARIMA(1,0,1)(1,1,1)
Rail Transport Services	none	ARIMA(0,0,3)(0,0,2) with mean
Cell Phones	1st and seasonal	ARIMA(1,1,0)(0,1,1)
IT Network Equipment	1st and seasonal	ARIMA(2,1,1)(1,1,1)
Printing Services	none	ARIMA(1,0,0)(2,0,0) with mean
Security Technology and Services	none	ARIMA(0,0,3)(0,0,2) with mean
Human Resources Management Services	1st	ARIMA(1,1,1)
Projection Techniques	1st	ARIMA(2,1,0)(2,0,0)
Restaurant Services	seasonal	ARIMA(0,0,2)(0,1,1) with drift
Facility Security Services	1st and seasonal	ARIMA(3,1,0)(1,1,0)

Furthermore, the column “Differencing” depicts whether the concerned subtotal was differenced to remove a trend or seasonality. This is equivalent to the d and D of the models. Most of the subtotals were differenced to remove either a trend or seasonality, or even both. In fact, only 4 models were fitted without any differencing, meaning that the corresponding subtotals are inherently stationary. The high need for differencing is not surprising as the products and services traded through framework agreements often depict autocorrelation between subsequent months and seasonal fluctuation, such as increased fuel consumption during winter months or reduced need for services during summer holidays, by nature, making the data nonstationary without differencing.

Only 7 out of the 26 models have a mean or a drift corresponding to δ in equation (5). This is rational as a constant mean of a nonstationary process reduces to a zero mean in a differenced process (Ruppert and Matteson, 2015). Thus, all the models without a mean or a drift have either 1st or seasonal differencing or both. However, 3 of the nonstationary subtotals have a mean with a linear deterministic trend, and thus, the models of these subtotals have differencing and a drift, i.e. a nonzero mean (see Ruppert and Matteson, 2015). Essentially a drift in a model reflects that the average of the volumes in these subtotals has increased over the years. Lastly, the 4 subtotals that are stationary by default, and hence need no differencing, all have a nonzero mean which is logical since trade volumes quite rarely have a zero mean because of the nature of the data.

All in all, the created models emphasise the strongly seasonal nature of the framework agreement volumes. Even though only half of the subtotals were seasonally differenced, 22 models have a seasonal part. Therefore, forecasting the future framework agreement volumes is in most cases supported by the volumes of the previous year. Moreover, the number of lags in all the models is at most 4 which is in line with the high amount of seasonal parts in the models. Fewer volumes from the previous months is needed when the year-old volumes from the corresponding months can be used for forecasting.

5.2 Forecast accuracy of the ARIMA models

The selection between the fitted models for each subtotal was based on the root mean squared error (RMSE) as it is generally recommended approach (Montgomery *et al.*, 2015). Hence, also the evaluation of the differences in the forecast accuracies between the chosen models and the current budgets was made with the RMSEs. The RMSE is defined as the square root of MSE in equation (B.33), and it describes the mean difference between the forecasted and the actual values of a time series. Thus, the smaller the RMSE, the better the model succeeds in forecasting the actual values.

The RMSEs of the budgets were calculated similarly to the models by determining the mean difference between the budgeted and the actual values. The out-of-sample data set containing the framework agreement trade volumes from January 2017 to October 2018 was used as the actual values in determining the RMSEs for the forecasts and the budgets of the same time period. In this case the RMSEs were acquired automatically with the R function *accuracy* (see Hyndman, 2018). In Table 7 are listed the RMSEs of both the chosen models and the current budgets.

Table 7. Comparison of the forecast accuracy of the chosen models and the current budgets with RMSE

Framework agreement subtotal	Model (EUR)	Budget (EUR)	Smallest RMSE
Occupational Health Care Services	328 270	413 739	model
Computers	975 311	901 676	budget
IT Consulting	873 978	1 393 757	model
Data Centre and Capacity Services	224 158	920 785	model
Leasing Services	4 075 707	3 825 803	budget
Cars	798 350	840 299	model
Microsoft	2 455 769	1 455 823	budget
Scheduled Flights	354 481	560 607	model
Facility Cleaning Services	188 835	1 536 253	model
Office Furniture	657 272	643 809	budget
Fuel	482 772	740 048	model
Domestic Accommodation	937 215	648 618	budget
Teleoperator Services	308 479	337 020	model
Management Consulting	313 499	554 546	model
Office and IT Supplies	142 178	228 835	model
Vehicle Leasing Services	469 262	573 898	model
Groceries and Non-food Products	77 046	101 548	model
Rail Transport Services	260 163	325 697	model
Cell Phones	162 608	501 046	model
IT Network Equipment	1 154 147	1 171 893	model
Printing Services	217 438	253 876	model
Security Technology and Services	651 423	863 845	model
Human Resources Management Services	190 449	172 857	budget
Projection Techniques	500 559	430 062	budget
Restaurant Services	426 586	357 535	budget
Facility Security Services	88 230	480 998	model

Comparison of the RMSEs shows that in 18 subtotals the model has better forecast accuracy than the budget. However, in 8 subtotals the current budget is predicting the actual volumes more accurately than the model as it has smaller RMSE. Ordered by the percentage difference between the RMSEs, these subtotals are:

1. Microsoft (68,69%)
2. Domestic Accommodation and Conference Services (44,49%)
3. Restaurant Services (19,31%)
4. Projection Techniques (16,39%)
5. Human Resources Management Services (10,18%)
6. Computers (8,17%)
7. Leasing Services (6,53%)
8. Office Furniture (2,09%)

As the RMSE is scale-dependent (Montgomery *et al.*, 2015), it cannot be used to evaluate the performance of the models compared to each other, only to the corresponding subtotal budgets. Also, the magnitude of the difference between the models and the budgets may be difficult to perceive from the RMSEs. Therefore, the forecast accuracies of the models and the budgets are also expressed as the percentage of MAE in equation (B.32), i.e. MAPE. For each subtotal the MAPEs of the model and the budget are listed in Table 8.

Table 8. Comparison of the forecast accuracy of the chosen models and the current budgets with MAPE

Framework agreement subtotal	Model	Budget	Smallest MAPE
Occupational Health Care Services	5.53%	7.11%	model
Computers	18.68%	15.39%	budget
IT Consulting	15.36%	23.65%	model
Data Centre and Capacity Services	5.09%	22.80%	model
Leasing Services	--	--	--
Cars	23.07%	27.36%	model
Microsoft	101.34%	64.29%	budget
Scheduled Flights	--	--	--
Facility Cleaning Services	4.72%	51.30%	model
Office Furniture	19.72%	20.81%	model
Fuel	15.94%	21.81%	model
Domestic Accommodation	--	--	--
Teleoperator Services	13.89%	15.09%	model
Management Consulting	24.70%	35.61%	model
Office and IT Supplies	10.36%	14.62%	model
Vehicle Leasing Services	62.18%	67.96%	model
Groceries and Non-food Products	5.16%	8.41%	model
Rail Transport Services	26.49%	23.32%	budget
Cell Phones	13.78%	46.10%	model
IT Network Equipment	57.60%	56.30%	budget
Printing Services	21.75%	27.54%	model
Security Technology and Services	65.70%	97.15%	model
Human Resources Management Services	22.93%	17.15%	budget
Projection Techniques	45.74%	79.11%	model
Restaurant Services	32.16%	31.37%	budget
Facility Security Services	11.23%	65.75%	model

The MAPEs give similar results about the forecast accuracies of the models compared to the budgets as the RMSEs. When measured with the MAPE the model is more accurate than the budget in 17 subtotals. The subtotals that have differences in the RMSE and the MAPE are Office Furnitures, Rail Transport Services, IT Network Equipment, and Projection Techniques. Also, for three subtotals – Leasing Services, Scheduled Flights, and Domestic Accommodation and Conference Services – the MAPE cannot be calculated because these subtotals have quarterly reporting, and thus, zero volumes in some months.

Even though there is some variation between the RMSEs and the MAPEs, it can be concluded that, in general, the ARIMA models reduce the magnitude of the forecast errors compared to the budget. Further, with some subtotals the reduction is outstanding, for example with Data Centre and Capacity Services or Facility Cleaning Services. Nevertheless, both the RMSE and the MAPE are only numerical representations of the forecast accuracy and tell nothing about how the models or the budgets are able to capture the trends or seasonality of the volumes. Hence, the graphical behaviour of the models and the current budgets must be considered in order to make reasonable conclusions about whether the ARIMA models improve the budgeting accuracy. Through a graphical inspection it can be also better understood why in the aforementioned 8 subtotals the budget is numerically more accurate than the model.

5.3 Effects of the ARIMA models on budgeting accuracy

To get an overall understanding of the effect the chosen models have on the budgeting accuracy, the behaviour of the forecasts is compared to the actual and the budgeted values graphically in addition to the RMSEs and the MAPEs. Based on the graphs and the numerical evaluation of the forecast accuracies the subtotals have been divided into five categories:

1. Model significantly better than budget.
2. Model better than budget.
3. Model slightly better than budget.
4. Model worse than budget.
5. Model incompetent.

Next, each category is addressed more specifically and the effect of the models on the budgeting accuracy is discussed within each category and in general.

5.3.1 Models more accurate than the current budgets

In 17 subtotals the ARIMA models are more accurate than the current budgets at forecasting the future framework agreement volumes. These subtotals comprise the first three categories in which the differences between the accuracy of the model and the budget range from significant to slight. The 5 subtotals of the first category in which the forecasts of the models are significantly better than the current budgets are depicted in Figure 5.

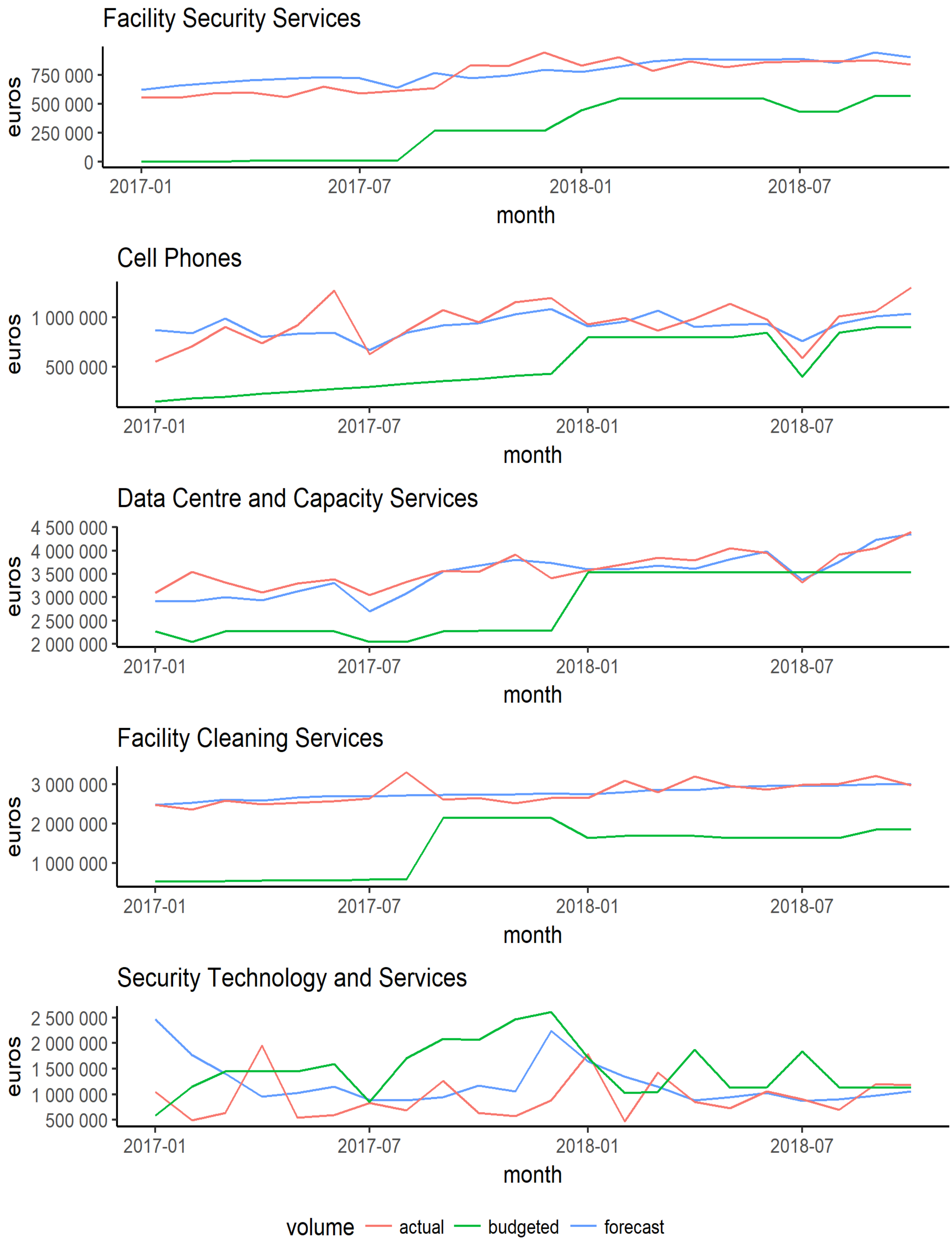


Figure 5. Subtotals in which the model is significantly more accurate than the current budget

In the first four subtotals the budgets have clearly been unable to predict the growth trend of the actual volumes. In Security Technology and Services, in turn, the budgeted volumes exceed the actual ones rather notably, especially in the second half of 2017. Even though the forecasts are not always able to capture the true behaviour of the framework agreement volumes in these subtotals, they are still in all cases considerably closer to the actual volumes compared to the budgets. Therefore, the dominance of the models in these subtotals can largely be explained with the glaring inaccuracy of the budgets. Consequently, using ARIMA models to forecast the future volumes of these 5 subtotals will certainly both enhance the budgeting process and increase the budgeting accuracy.

The second category consists of 6 subtotals in which the superiority of the models is not as apparent as in the first category. For these subtotals, depicted in Figure 6, the budgets are clearly more accurate than in the previous category. However, the models are better at forecasting the magnitude of the fluctuations of the actual volumes. In the first five subtotals this can be at least partly explained with the seasonality of the subtotals as seasonal ARIMA models are specifically fitted to detect seasonal fluctuations, even the ones that are not so obvious to the naked eye. Therefore, even though the ARIMA models are unable to capture the exact behaviour of the framework agreement volumes, the budgeting accuracy of the subtotals of also this category can be improved with the models.

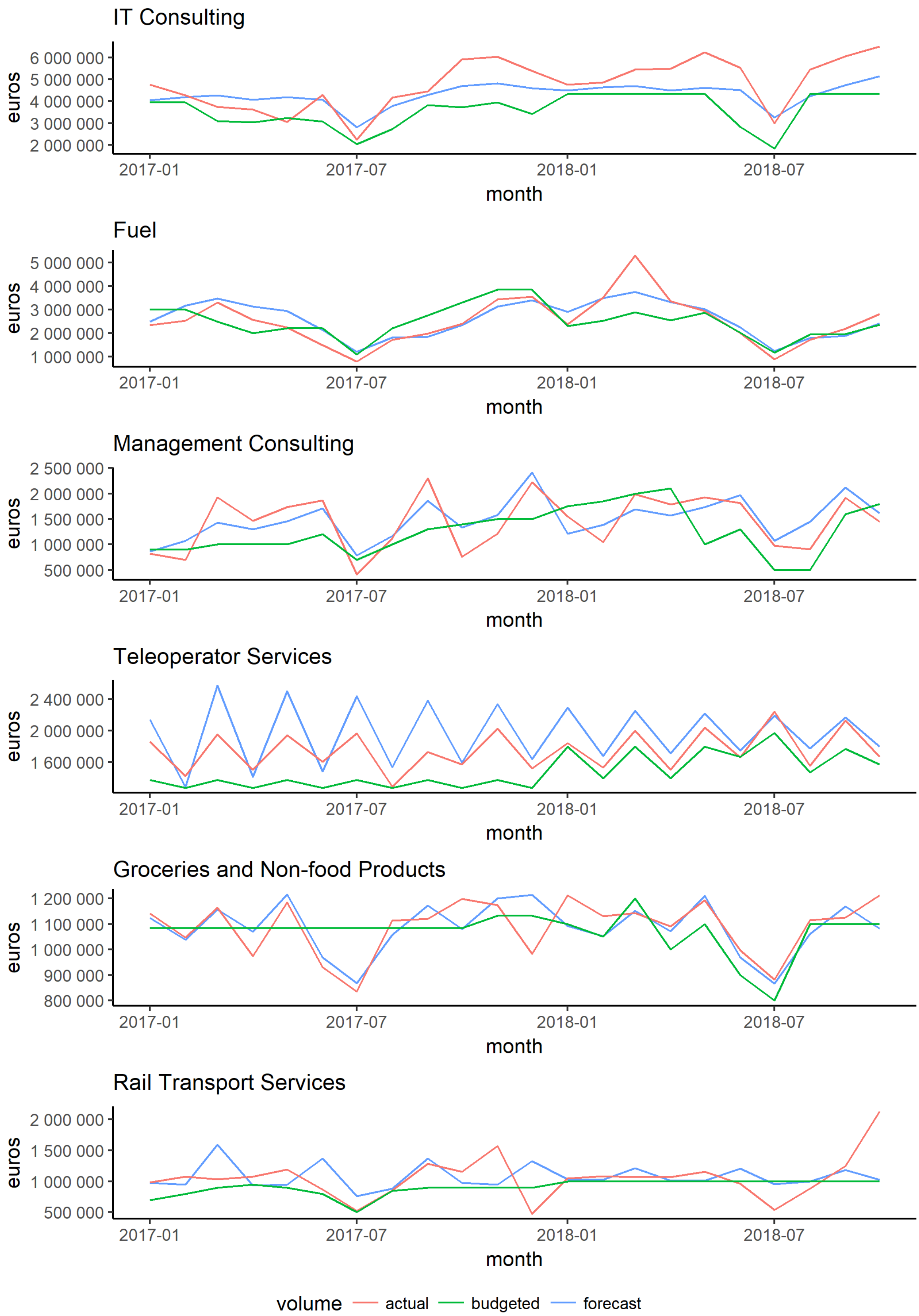


Figure 6. Subtotals in which the model is more accurate than the current budget

Lastly, Figure 7 shows the third category of subtotals in which the models perform better than the budgets. In these 6 subtotals only slight difference between the forecasts of the models and the budgets can be detected. 5 of the subtotals have distinct seasonality which supports the good performance of the models. In turn, the high accuracy of the budgets, especially in Occupational Health Care Services and Scheduled Flights, probably stems from the long history of these subtotals. In fact, all the subtotals in this category have existed for over 10 years. In addition, the products and services of these framework agreements are needed constantly and purchased quite regularly. Hence, the budgeting of these framework agreement volumes could have been developed to a rather accurate level. IT Network Equipment is an exception when compared to the other subtotals in this category. It has also been in Hansel's framework agreement range since 2007, but for some reason the volumes budgeted for it are almost constant throughout the observation period. Therefore, as the model forecasts more fluctuation in the volumes, it performs overall better than the budget, even though it cannot forecast the great spike in June 2018. All in all, since all the models in this category have smaller RMSE than the budgets, the models can further increase the accuracy of budgeting these subtotals. Hence, using ARIMA models during the budgeting process is beneficial for also these subtotals.

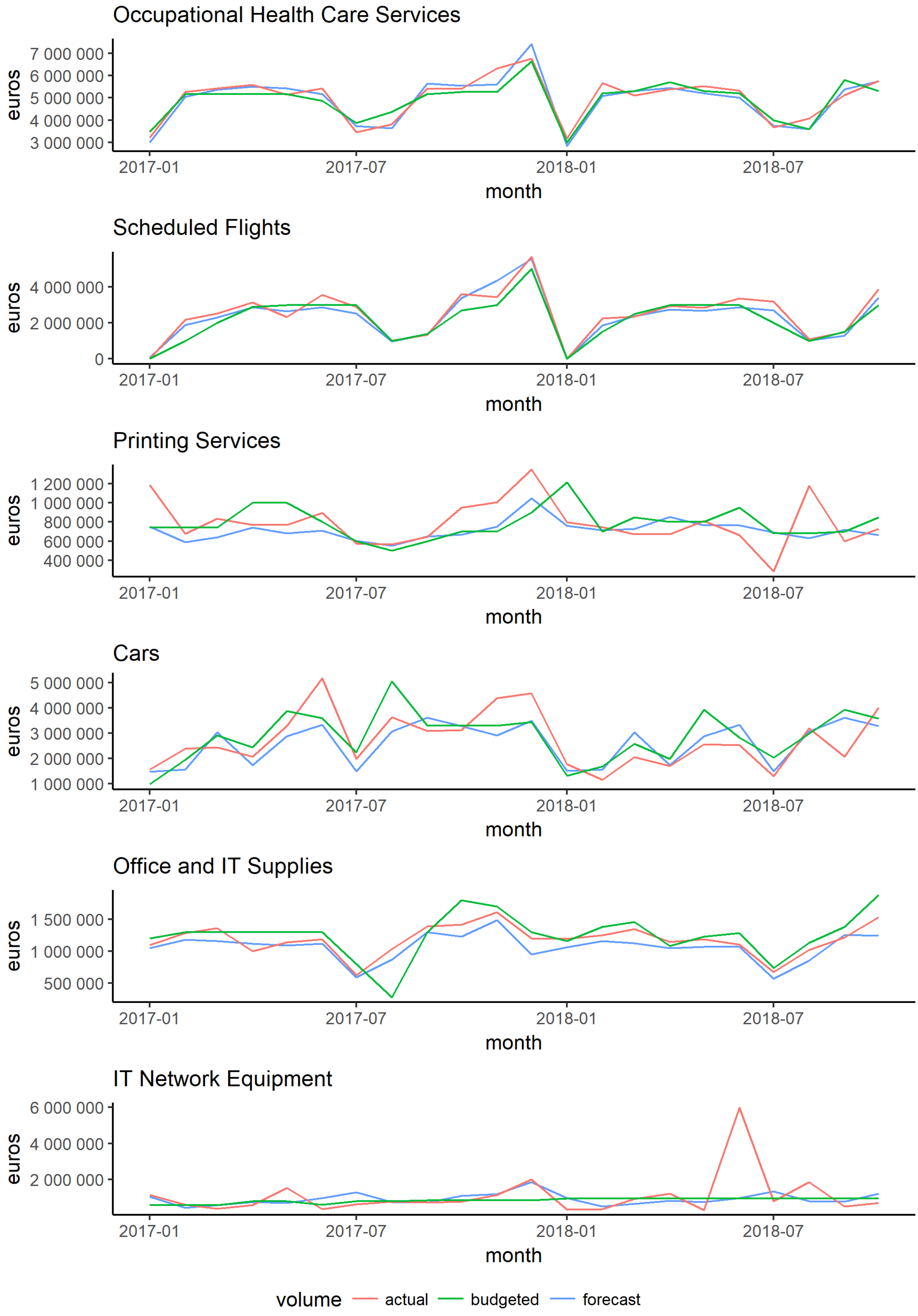


Figure 7. Subtotals in which the model is slightly more accurate than the current budget

5.3.2 Models less accurate than the current budgets

The fourth category includes 6 subtotals in which the budgets are more accurate at forecasting the future framework agreement volumes than the models when measured with the RMSE. However, for two subtotals of this category – Office Furniture and Projection Techniques – the model has smaller MAPE than the budget. This emphasises the problem of numerical accuracy measurements as they can give contradictory results and penalise individual forecast errors disproportionately. Hence, even though the budgets of these subtotals have smaller RMSEs than the models, it does not necessarily mean that the models are incapable to depict the fluctuations of the actual volumes.

The ability of the models to capture the behaviour of the framework agreement volumes can be seen from the graphs in Figure 8. For example, in Domestic Accommodation and Conference Services the model forecasts the actual framework agreement volumes as well as the budget, and sometimes even better, but fails to take into account the regularly occurring months without any framework agreement trade caused by quarterly reportage. This is an exemplary instance of a subtotal for which the budgeting accuracy can be improved by combining the forecasts of the model with the in-house tacit knowledge about the framework agreement behaviour. Another distortion of numerical forecast accuracy can be detected in Microsoft. The model forecasts the actual volumes almost flawlessly but the forecast for January 2018 is not high enough causing both the RMSE and the MAPE to indicate bad forecast accuracy compared to the budget. Therefore, although the budgets of these subtotals have smaller RMSEs than the models, the budgeting accuracy can still be enhanced by using the models together with tacit knowledge. Moreover, utilising ARIMA models during the budgeting will also solve other deficiencies of the process, such as providing a fact base for the budget instead of subjective intuition.

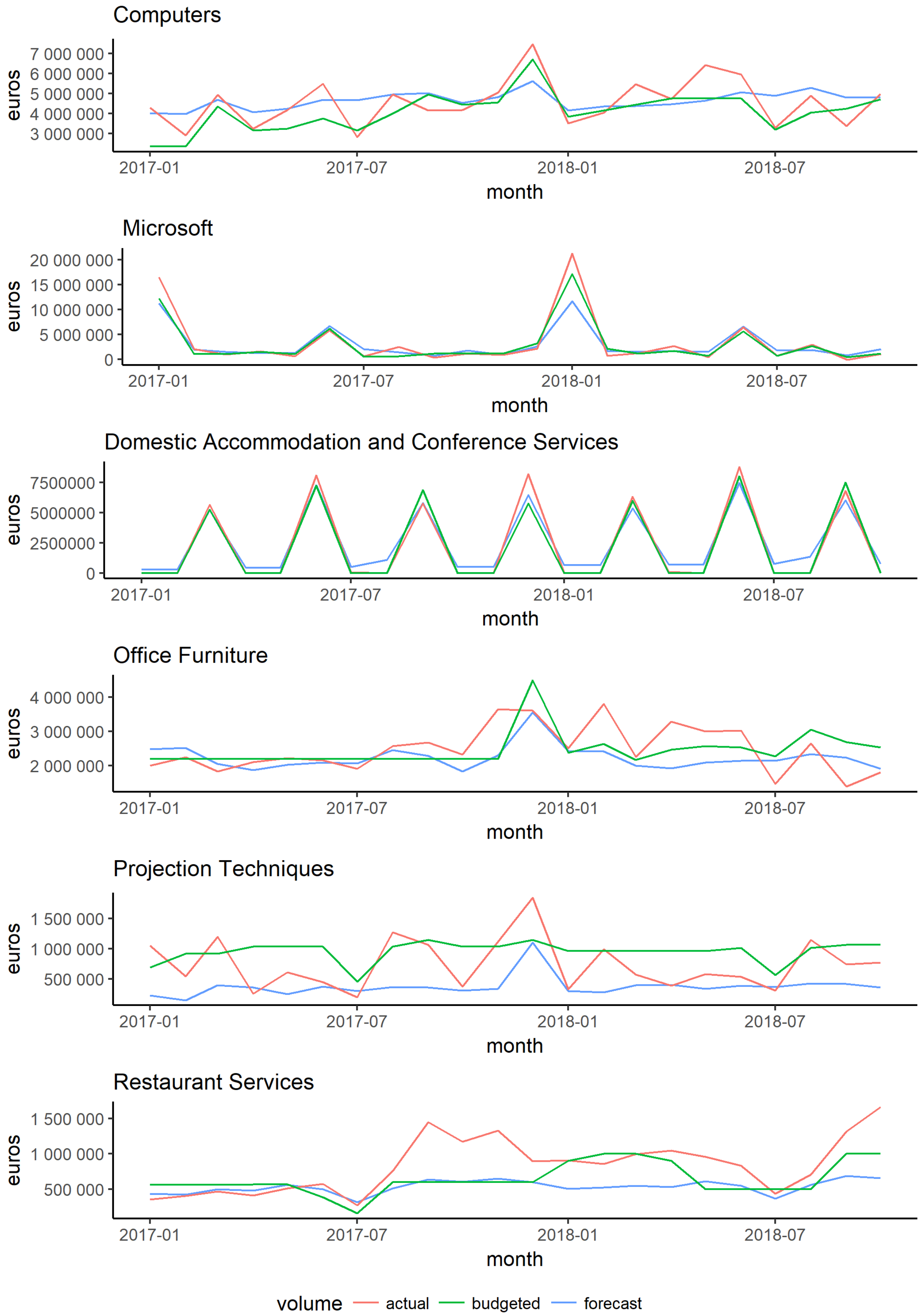


Figure 8. Subtotals in which the model is less accurate than the current budget

5.3.3 Models incompetent to forecast the framework agreement volumes

In the last category, presented in Figure 9, are the 3 subtotals for which the models were found to be incompetent to produce usable forecasts. For Leasing Services and Human Resources Management Services the model forecasts practically a constant volume for each month. The forecast for Vehicle Leasing Services, in turn, has some fluctuation but it is simply too minor to be relevant when budgeting the actual framework agreement volumes.

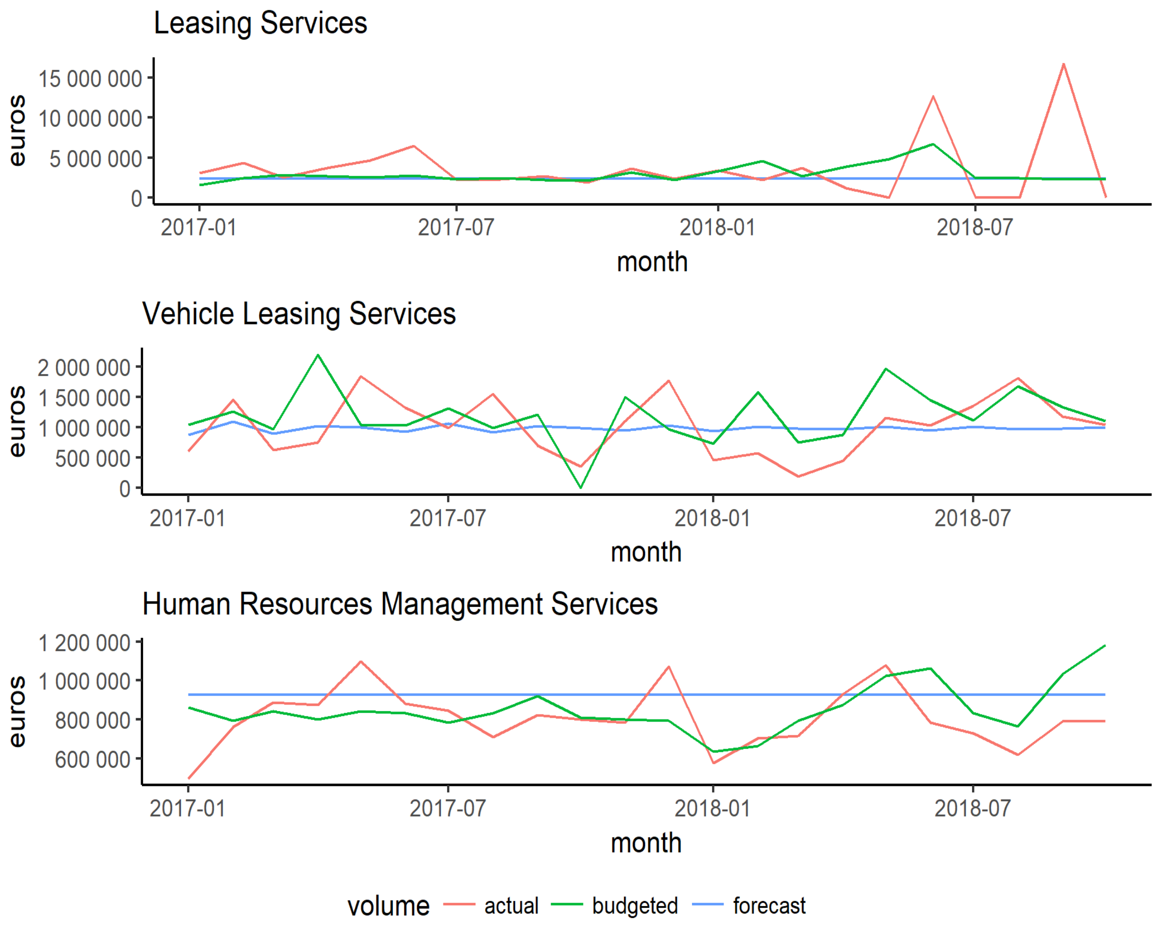


Figure 9. Subtotals in which the model is incompetent to forecast the framework agreement volumes

In Human Resources Management Services, the incompetence of the model may be caused by a lack of observations. The framework agreements included in the subtotal have only begun in February 2015 resulting in 45 observations in total. Therefore, it is not unusual that the model is unable to capture the true behaviour of the framework agreement volumes and produce accurate forecasts for the next 22 months when fitted to an in-sample data set with only 23 observations. However, the underperformance of the models in Leasing

Services and Vehicle Leasing Services cannot be explained with too small in-sample data sets as both the subtotals have data from 2007 onwards.

Consequently, the subtotals of this category have such behaviour that cannot be captured by the models created in this study. The models created for these subtotals all have only non-seasonal parts. Disregarding the seasonal parts may be causing the poor performance of the models as at least Leasing Services and Vehicle Leasing Services should have some seasonality caused by the expiration of the lease contracts. Hence, further research would be needed to inspect whether other time series models could forecast the concerned subtotals better and if gaining more observations for Human Resources Management Services could improve the performance of the model. Concluding, when concerning these three subtotals, the accuracy of budgeting may not be improved with the ARIMA models created in this study.

5.4 Summary of the results

The ARIMA models chosen for each subtotal have been introduced and their forecast accuracies compared to the current budgets has been evaluated with both numerical and graphical criteria. Now the results are reflected in the light of the research questions, and their significance to Hansel, other public procurement units, and future public procurement research in general are considered.

5.4.1 Ability of the ARIMA models to forecast framework agreement volumes

This thesis has examined the capability of ARIMA time series models to capture the behaviour of framework agreement volumes and through that their usability for forecasting governmental procurement volumes. The results cover several gaps in previous literature and provide practical contribution also beyond Hansel’s objectives. As stated, public procurement is underrepresented in previous research compared to private procurement despite its central role in delivering public services (Knight *et al.*, 2012; Flynn and Davis, 2014). Therefore, conducting a study which addresses the central government’s centralised procurement is a contribution to public procurement research in itself.

Furthermore, there is a need for greater theoretical rigour in public procurement research, since only 29 percent of public procurement articles are theoretically grounded (Flynn and Davis, 2014). Even though this thesis does not develop a theoretical framework, it is based on the theory of ARIMA time series models and how they can be used to support decision making. Time series models have previously been used to approach phenomena related to procurement, and even public procurement, as shown in Chapter 3. However, the framework agreement volumes of governmental procurement have not been modelled with such models. Hence, the created models offer a new approach for public procurement research and answer to the called for utilisation of descriptive, predictive, and prescriptive modelling in procurement literature (Matopoulos *et al.*, 2016).

Only 3 of the 26 models were incompetent to forecast the actual volumes of the framework agreements. Conversely, the rest of the models were able to capture the behaviour of the volumes either approximately (e.g. Cell Phones), with few missteps (e.g. Scheduled Flights), or almost precisely (e.g. Microsoft). Consequently, it can be stated that ARIMA time series models are in most cases capable of detecting the characteristic behaviour of framework agreement volumes and based on these features producing accurate and reasonable forecasts of the future volumes.

Seasonal fluctuation is apparent in most of the subtotals; as much as 22 models had seasonal terms to capture the seasonality of the volumes. 14 subtotals have such strong seasonality that they needed to be seasonally differenced before model fitting, as for the rest subtotals adding seasonal AR or MA terms was enough for the model to be able to capture the fluctuations correctly. Strong presence of seasonality is natural when considering the nature of the products and the services acquired through framework agreements. For instance, cell phones and cars have a certain life cycle after which new ones are purchased. Moreover, software licenses as well as products and services that are leased as licenses or lease contracts have a duration for the contract after which it needs to be renewed. This causes regular spikes to the volumes of the framework agreements used for such purchases.

As the ARIMA models were found to be competent to forecast framework agreement volumes accurately, this thesis has practical implications also outside Hansel. In recent years many public and private organisations have centralised their purchasing with framework agreements which have established a key role at least in the EU public tendering (Lempinen, 2013; Andrecka, 2016). Nonetheless, generating realistic forecasts of framework agreement throughputs have often proven to be problematic (Andrecka, 2016).

The models created in this thesis cannot give doubtless forecasts of future framework agreement volumes but combined with the tacit knowledge of procuring professionals they can decrease the uncertainty of the prediction process. In addition, using fact based and provably effective time series models as the foundation for the forecasts instead of mere intuition improves the creditability of the volume predictions and the whole organisation. Therefore, even though the effects of the models created here are not directly generalisable, also other public procurement units at least in Finland and in Europe should explore the positive effects time series models could have on their forecasts of future framework agreement volumes.

5.4.2 Ability of the ARIMA models to improve budgeting accuracy

The impact of the created ARIMA models on Hansel’s current budgeting process and whether the models will improve the budgeting accuracy wrap up the empirical contribution of this thesis. For Hansel, the need to explore if time series models could enhance the current budgeting process derives from the need to better achieve its zero profit objective through more accurate allocation of the framework agreement service fees. As there are currently rather vast differences between the actual volumes and the framework agreement and customer specific budgets, improving the budgeting accuracy would enhance achieving the zero profit objective.

When compared to the current budgets, most of the models performed better at forecasting the actual framework agreement volumes. If the forecasting accuracy is evaluated exclusively with the RMSEs, the forecasts of 18 models were more accurate than the corresponding budgets. When evaluated with the MAPEs, the models outperformed the budgets in 17 subtotals. However, the numerical measurements of forecast accuracy do not account for how well the models predicted the fluctuations of the volumes. Based on both the RMSEs, the MAPEs, and the graphical accuracies of detecting the fluctuations, 17 models were more accurate than the budgets. Also, for 4 of the 6 subtotals with the budgets having smaller RMSEs, the differences in the graphical forecast accuracies between the models and the budgets was mostly trivial. Therefore, it can be concluded that in as much as 23 subtotals the usage of ARIMA models will improve the budgeting accuracy. All in all, utilising the created models in Hansel’s budgeting would result in a 6.34 percentage points, or EUR 76 million, more accurate budget than with the current process.

Nevertheless, using the models alone could reduce the overall budgeting accuracy from what it is now. The current process relies strongly on the knowledge of the category and the account managers of the procurement needs of their customers. Discarding this insider information would prevent correctly budgeting anomalies in the framework agreement trade which the models cannot forecast on the basis of the historical volumes. Hence, the models should be implemented as a part of the budgeting process so that the forecasts generated by them would be adjusted according to the insights of the category and the account managers if needed. At the same time, the category and the account managers would be more comfortable with the initial budget as they knew it was generated by generally acknowledged mathematical models rather than the top management's intuition about and expectations for the volumes of the coming year.

6 Conclusions

The purpose of this thesis has been to find means for the case company to better achieve zero profit by enhancing the decision making of its board regarding the revenue. As the revenue depends mainly on the volumes of framework agreement trade, budgeting the future volumes accurately is essential for effective decision making. Thus, this thesis has aspired to transform the knowledge generated through research into a decision making tool for a real life business problem after Flynn and Davis (2014). At the same time, an understudied approach to public procurement has been addressed by utilising time series models to forecast the future framework agreement volumes. In this chapter the research, its main results, and the managerial implications based on them are summarised. Also, the limitations of the study are recognised and suggestions for future research are given.

6.1 Research summary

This thesis has combined forecasting governmental procurement with time series models, enhancing budgeting with broad utilisation of data, and supporting decision making with predictive analytics through a case company. Hansel is a non-profit Finnish central purchasing body (CPB) which tenders the centralised procurement of the central government through framework agreements according to the national and the European public procurement legislation (see 1096/2008, Directive 2014/24/EU). Being a CPB, Hansel operates in the field of public procurement which despite its societal and economical importance is lacking behind private procurement in research (see Thai, 2001; Knight *et al.*, 2012).

Around 87 percent of Hansel's revenue is comprised of service fees from the framework agreement trade. The volumes of the coming year's trade are budgeted on a framework agreement and a customer level, and the service fees are determined based on the concerned budgets. As Hansel has a zero profit objective, setting the right level for the service fees is a central, even a strategic, decision for the board. Unfortunately, achieving zero profit has been difficult for Hansel because of inaccuracies in the budget. On average the upper-level budget has differed from the actual total around 5 percent, but the differences between the budgeted and the actual volumes on both the framework agreement and the customer level have been significant during the past years, ranging from few to even 100 percent. In addition to inaccuracy, the current budgeting process has also other inefficiencies, including repetitive tasks, strong emphasis on the intuition of the top management, ostensible participation, and lacking utilisation of data.

To improve the accuracy of Hansel's current budgeting process and to provide new insight into the usage of time series analytics in public procurement context two research questions were formed. The first research question addresses the general contribution of time series models to public procurement research by assessing the ability of the models to capture the behaviour of the framework agreement volumes. The second question, in turn, concerns the effect of the models on the budgeting accuracy compared to Hansel's current budgeting process. Time series models were chosen for forecasting because the monthly framework agreement volume data represents a time series with its equally spaced observations (Box *et al.*, 2008), and since time series models are efficient in capturing the dependent structure of time series (Montgomery *et al.*, 2015). Further, autoregressive integrated moving average (ARIMA) models were used because they have a finite, preferably small, number on parameters (Ruppert and Matteson, 2015) which makes model fitting and forecasting fast, straightforward, and rather accurate.

The framework agreement volumes were handled on a subtotal level in which the volume of framework agreements with similar product and service categories are combined into one subtotal. That way a broader portion of the framework agreements could be modelled, the problem of overlapping framework agreement contract periods was eliminated, and the number of models was kept manageable. The selection of the modelled subtotals was based on their importance to Hansel's overall revenue; all subtotals with total volume over EUR 10 million in 2017 were chosen. Hence, around 85 percent of the framework agreement trade in 2017 was covered with around a third of all subtotals. Some additional modifications were made based on the insights and wishes of the CCO and the CFO so that eventually 26 framework agreement subtotals were modelled with ARIMA models.

The model fitting and the evaluation of forecast accuracy was done in R. Each subtotal data was divided into an in-sample and an out-of-sample data set to get an unbiased assessment of the forecast accuracies. Based on the plotted subtotals and their autocorrelation (ACF) and partial autocorrelation (PACF) functions about ten models on average were fitted to the in-sample data sets both manually and automatically with the R functions *Arima* and *auto.arima*, respectively (see Hyndman, 2018). Similarly, the forecasts and the forecast accuracy measurements were calculated with the R functions *forecast* and *accuracy*, respectively (see Hyndman, 2018). The best model for each subtotal was chosen based on the smallest root mean squared error (RMSE) as in general it results in the best fit to the time series (see Montgomery *et al.*, 2015). Finally, the forecasting accuracy of the chosen models and the current budgeting process was compared both numerically and graphically to get an overall understanding of the ability of the models to capture the framework agreement behaviour and to improve Hansel's budgeting accuracy.

Based on the forecast accuracy of the created models and their ability to detect the behaviour of the framework agreements the results of this thesis support the usage of ARIMA models in forecasting framework agreement volumes. Only three of the models were unable to forecast the seasonal fluctuations and other characteristic behaviour of the framework agreement volumes. Moreover, 17 of the competent models performed better than the corresponding budgets both numerically and graphically producing on average 36.6 percent smaller forecast errors when measured with the RMSEs. In addition, even though for six subtotals the budgets had smaller RMSEs than the models, the models still captured the general behaviour of the subtotals very closely and were only at times more inaccurate than the budgets. Therefore, concerning the first research question, in almost 90 percent of the addressed subtotals the ARIMA models were able to capture the behaviour of the framework agreement volumes in detail.

Also, the effect of the models on Hansel's budgeting accuracy is notable. Across the modelled subtotals the forecasts of the models were 6.34 percentage points more accurate than the budget which equals to around EUR 76 million. Hence, regarding the second research question, the models will increase Hansel's budgeting accuracy compared to the current budgeting process. Further, through improving the budgeting accuracy, the models will also improve the correct alignment of service fees for achieving zero profit. However, the models cannot predict anomalies such as new customers or changes in the framework agreements. Moreover, in three subtotals the models were unable to predict the months with zero trade caused by divergent reporting. Therefore, the future budgeting should not be based solely on the models, but the forecasts need to be adjusted to factor in special features known by the customer and the category managers and the controllers. By combining the forecasts with the in-house tacit knowledge, the most benefit from the ARIMA models can be achieved in the budgeting process.

6.2 Managerial implications

The results of this research indicate that the ARIMA models will improve Hansel's budgeting accuracy compared to the current budgeting process. Therefore, there are clear managerial recommendations for Hansel as the utilisation of the models is strongly supported. However, not all decisions should be grounded entirely on analytical models, and this applies also to Hansel's budgeting process. Framework agreement volumes are exposed to anomalies differing from the long-term behaviour, which the models are unable to forecast. Hence, the forecasts generated by the models should be supported and adjusted by the valuable tacit knowledge of the category and the account managers about the procurement needs and behaviour of their customers.

Since the models can address several deficiencies of the current budgeting process, such as inaccuracy, repetitive manual labour, and budgets based on subjective opinions, they should be implemented as a part of the whole process. In fact, the budgeting process should be built on the forecasts of the models to provide a rigorous fact base for the framework agreement budgets. The following steps should therefore be made to utilise the models during the budgeting process:

1. Forecasting the future framework agreement volumes monthly with an R script.
2. Collecting the progress of the actualised and the forecasted future volumes in a QlikView report in real-time for everyone to inspect.
3. Using the forecasted volumes as an initial budget and adjusting them according to the tacit knowledge of the category and the account managers.

In practice, the models are best implemented through the analysts. The new, suggested budgeting process, depicted in Figure 10, starts with the forecasts of the models being used as a starting point for the whole budgeting. Throughout the year the models generate forecasts for the subtotal volumes monthly for the next 18 months based on the data available for the actual volumes up to that point. Setting the forecast horizon to 18 months assures that there are forecasts up to the end of the next year when the budgeting process starts in the autumn. By revising the forecasts every month, they are as up-to-date as can be. The forecasts and the actual framework agreement volumes are collected in a report in QlikView, a business intelligence software used by Hansel, so that all employees can follow their progress on a monthly basis. Therefore, everyone can inspect the forecasts on which the upper-level budget is based, which makes the upper-level budget more justifiable and increases the level of transparency in setting it. This has been called for especially by the category and the account managers.

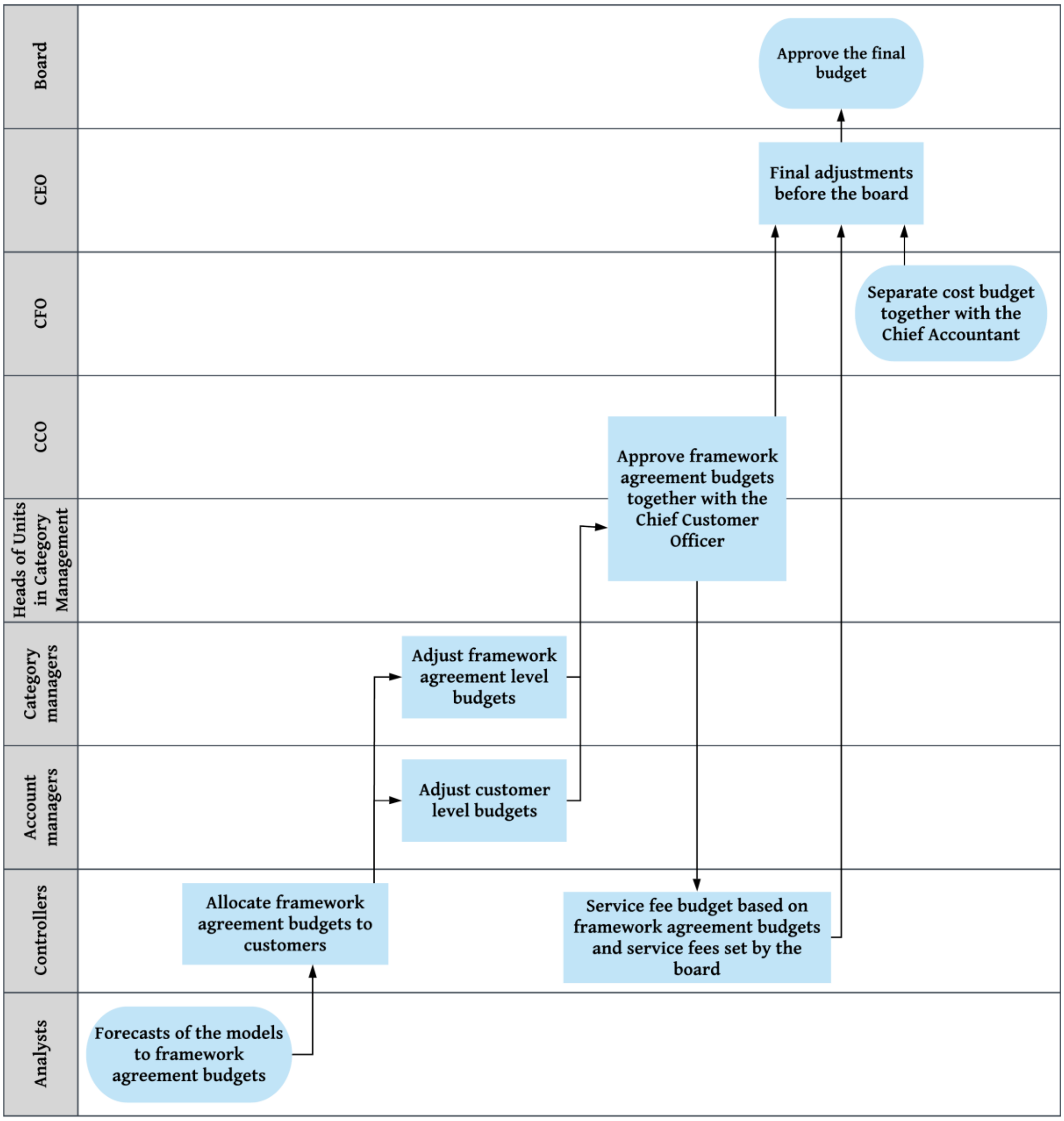


Figure 10. Suggested budgeting process with the ARIMA models implemented as a part of it

As the actual budgeting begins in the autumn, the analysts assemble the forecasted volumes for the coming year as the framework agreement level budgets. Next, these budgets are allocated to the customers by the controllers based on the customer distributions of the previous years. Then both the framework agreement and the customer level budgets are adjusted by the category and the account managers, respectively, according to their tacit knowledge. Lastly, the budgets are approved together by the CCO, the Chief Customer Officer, and the Heads of Units in Category Management before the final approvals of the CEO and the board. Separate budgets for the service fees and costs are formed similarly as currently.

Based on the results, 23 of the models can be implemented right away. Implementing the three incompetent models, in turn, should be further considered, and additional inspection could be conducted to find out why the models failed to capture the behaviour of the concerned subtotals. Nevertheless, also the incompetent models could be used to give a starting point for the budgeting as the forecasted volumes represented most likely the long-term mean without any fluctuation. In that case, the forecasts of these models would merely need to be manually adjusted more heavily than with the other modelled subtotals.

Compared to the current budgeting process depicted in Figure 2 utilising the ARIMA models in the new budgeting process has several benefits. Using the models as a base for the budgeting will increase the level of automation and the efficiency of the process as well as prune repetitive stages. In addition, the models will enhance the transparency of the process by providing a budget based on facts instead of intuition. Finally, overlooking the insights of the category and the account managers will be reduced since the number of stages in which the budget is adjusted by the CCO and the CEO according to their intuition is decreased.

Concluding, the potential benefits for Hansel gained from utilising the ARIMA models during the budgeting process are:

1. Overall budgeting accuracy is increased around 6 percentage points.
2. Subtotal level inaccuracies are decreased on average 37 percent.
3. Steps required in the budgeting process are reduced by one.
4. Transparency of the budgeting is improved with a fact base.
5. Adjusting suitable service fees is eased along with the increased budgeting accuracy.

To achieve the aforementioned benefits the current budgeting process needs to be revised before implementing the models. Since the CEO and the CCO have gotten used to setting more of a sales target for the framework agreement volumes than an actual prediction of the future volumes, setting the upper-level budget according to the forecasts may take some accustoming. Hence, the roles and responsibilities of everyone involved in the budgeting process need to be carefully clarified and redefined before implementing the models into the process to fully benefit from them and better achieve the zero profit objective.

6.3 Limitations

Some limitations need to be considered when assessing the results of this thesis. First, this study does not cover the model deployment phase, and therefore, the actual effects of the models on the budgeting accuracy and achieving zero profit are not observed. Hence, the results about the performance of the models compared to the current budget are only theoretically justified without applied knowledge. In order to detect the actual effects of the models on the budgeting process and adjusting the service fees, Hansel's budgeting accuracy needs to be assessed after the models have been implemented.

Second, the modelling and determining the forecast accuracies of the models were done with ready-made R functions. Popular R packages stored in the Comprehensive R Archive Network (CRAN), such as the used *forecast*, are generally approved to be reliable and produce accurate results because of the strict policy and checks CRAN has for the packages (Claes *et al.*, 2014). Nevertheless, there is always a possibility of human error or technical deficiency. In addition, no comparison between different R packages was made to determine if another package would have been better suited for the purposes of this study.

Third, even if the models could forecast the future framework agreement volumes perfectly, they cannot eliminate the anomalies caused by Hansel's customers and suppliers. There are shortcomings in the procurement planning of the customers and their procurement behaviour is somewhat irrational, which affect the final volumes of the products and services they acquire through framework agreements. The models are unable to predict such irregular changes to the framework agreement volumes, and the account managers cannot adjust the forecasts to match the reality, if the customers themselves do not know their procurement needs beforehand for the coming year. Moreover, as the customers are a part of the central government, their procurement budgets and resources are directly dependent on the Finnish government budget and the political decisions affecting it. These are clearly out of the forecasting capabilities of the models. In addition, the data Hansel has on the framework agreement volumes rests on the trade reported by the suppliers. If there are errors or missing trade in the reporting, the amount of the actual volumes is distorted which naturally affects the forecasting accuracy of the models.

Finally, Hansel is facing some changes which will affect predicting the future framework agreement volumes. The models created here are based on the volumes of the central government's procurement. Hence, they most likely need to be updated after also municipalities become Hansel's customers along with the upcoming merger. Moreover, Hansel has put to use dynamic purchasing systems (DPSs) in four framework agreements and is planning more. In DPSs, the structure and the volume of purchases is determined by the needs of the customers in the boundaries of the framework agreements. Hence, the volume predictions will differ from traditional framework agreement trade and budgeting them cannot be made on the same basis as the budgets of the traditional framework agreement volumes. Hence, the created models do not take into account the effect of DPSs on the future framework agreement trade.

6.4 Suggestions for future research

This thesis is the first known study approaching the forecasting of the central government's centralised procurement with time series models. Hence, there are numerous avenues for future research to continue from. First, only a portion of Hansel's framework agreement subtotals were modelled. Therefore, comprehensive conclusions about the ability of the models to capture the behaviour of all subtotals cannot be made. Also, the subtotals include multiple framework agreements. Hence, to get a better understanding of the performance of the models in accurately forecasting the framework agreement volumes and the impact on Hansel's budgeting process, the models should be fitted on individual framework agreements instead of subtotals.

Second, assessing the ability of the models to improve Hansel's current budgeting process was based on the numerical and graphical differences of the forecasts and the budgets compared to the actual volumes for the last two years of the used data. Further research is needed to evaluate the effect of the models on the budgeting accuracy and through that the correct adjustment of the service fees after the models have actually been implemented to the budgeting process. That way also the last steps of the forecasting process by Montgomery *et al.* (2015) would be covered. By extending the research period also the effects of the merger and the introduction of DPSs on achieving zero profit could be assessed.

Third, this thesis concerned only traditional ARIMA models with the assumption that future framework agreement volumes can be fully explained with the historical volumes. In reality, also many external things may have an effect on the framework agreement trade. As stated in the limitations, the trade is dependent on the procurement needs of the customers, which in turn depend on the state budget. In addition, the behaviour of a certain framework agreement might be explained by the volumes of other framework agreements, for instance the volumes of cars and car insurances most likely have some dependency with each other. Also, social and environmental events and changes, such as the refugee crisis or a cold winter, may affect the volumes of specific framework agreements. Hence, forecasting framework agreement volumes with time series models such as ARIMAX or VAR, which consider also the effect of explanatory variables, could give a better understanding of the behaviour of the volumes and result in even more accurate forecasts than the models created here.

Lastly, this thesis is limited to concern only the Finnish central government and its centralised procurement. Hence, the results are not straightforwardly generalisable to other countries or even to the framework agreements used by Finnish municipalities and joint municipal authorities. However, also other framework agreements and public procurement contracts most likely have similar characteristic behaviour as observed here that can be accurately captured by time series models. As framework agreements have established a key role in EU public procurement, replicating this research in other European countries would give a broader understanding of the benefits of utilising time series models in forecasting public procurement volumes.

References

1096/2008 *Laki Hansel Oy -nimisestä osakeyhtiöstä*. Available at: https://www.finlex.fi/fi/laki/alkup/2008/20081096 [Accessed: 17 October 2018].

Acharya, R. N., Kagan, A., and Manfredo, M. R. (2009) Impact of Rising Fuel Cost on Perishable Product Procurement. *Journal of Business Logistics*, 30(1), pp. 223–242. Wiley Online Library.

Albano, G. L. and Sparro, M. (2010) Flexible Strategies for Centralized Public Procurement. *Review of Economics and Institutions*, 1(2), pp. 1–32.

Andrecka, M. (2016) Dealing with Legal Loopholes and Uncertainties within EU Public Procurement Law Regarding Framework Agreements. *Journal of Public Procurement*, 16(4), pp. 505–527. Emerald Publishing.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2008) *Time Series Analysis: Forecasting and Control*. 4th ed. John Wiley & Sons.

Brammer, S. and Walker, H. (2011) Sustainable Procurement in the Public Sector: An International Comparative Study. *International Journal of Operations & Production Management*, 31(4), pp. 452–476. Emerald Publishing.

Chandra, S. R. and Al-Deek, H. (2009) Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. *Journal of Intelligent Transportation Systems*, 13(2), pp. 53–72. Taylor & Francis.

Cheung, C. F., Lee, W. B., Wang, W. M., Chu, K. F., and To, S. (2003) A Multi-Perspective Knowledge-Based System for Customer Service Management. *Expert Systems with Applications*, 24(4), pp. 457–470. Elsevier.

Claes, M., Mens, T., and Grosjean, P. (2014) On the Maintainability of CRAN Packages. *Software Evolution Week – IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE)*, pp. 308-312. IEEE.

Confederation of Finnish Industries (2018) *Julkiset hankinnat*. Available at: https://ek.fi/mita-teemme/yrityslainsaadanto/julkiset-hankinnat/ [Accessed: 25 October 2018].

Davenport, T. H. (2006) Competing on Analytics. *Harvard Business Review*, 84(1), pp. 98–108.

Davenport, T. H. (2009) Make Better Decisions. *Harvard Business Review*, 87(11), pp. 117–123.

Davenport, T. H., Harris, J. G., De Long, D. W., and Jacobson, A. L. (2001) Data to Knowledge to Results: Building an Analytic Capability. *California Management Review*, 43(2), pp. 117–138. SAGE Publications.

Dekimpe, M. G. and Hanssens, D. M. (2000) Time-Series Models in Marketing: Past, Present and Future. *International Journal of Research in Marketing*, 17(2-3), pp. 183–193. Elsevier.

Dykes, B. (2017) *Crawl with Analytics Before Running with Artificial Intelligence*. Available at: https://www.meltwater.com/nz/blog/crawl-analytics-running-artificial-intelligence/ [Accessed: 22 November 2018].

Ediger, V. Ş. and Akar, S. (2007) ARIMA Forecasting of Primary Energy Demand by Fuel in Turkey. *Energy Policy*, 35(3), pp. 1701–1708. Elsevier.

Erridge, A. and McIlroy, J. (2002) Public Procurement and Supply Management Strategies. *Public Policy and Administration*, 17(1), pp. 52–71. SAGE Publications.

European Commission (2015) *Single Market Scoreboard. Performance per Policy Area: Public Procurement*. Available at: http://ec.europa.eu/internal_market/scoreboard/_docs/2015/09/public-procurement/2015-09-scoreboard-public-procurement_en.pdf [Accessed: 7 November 2018].

European Commission (2018) *Public Procurement*. Available at: https://ec.europa.eu/growth/single-market/public-procurement_en [Accessed: 23 October 2018].

European Parliament and of the Council (2014) *Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on Public Procurement and Repealing Directive 2004/18/EC (OJ L 94).* Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014L0024 [Accessed: 23 October 2018].

Flynn, A. (2018) Measuring Procurement Performance in Europe. *Journal of Public Procurement*, 18(1), pp. 2–13. Emerald Publishing.

Flynn, A. and Davis, P. (2014) Theory in Public Procurement Research. *Journal of Public Procurement*, 14(2), pp. 139–180. Emerald Publishing.

Hahn, H., Meyer-Nieberg, S. and Pickl, S. (2009) Electric Load Forecasting Methods: Tools for Decision Making. *European Journal of Operational Research*, 199(3), pp. 902–907. Elsevier.

Halonen, K.-M. (2016) Disclosure Rules in EU Public Procurement: Balancing between Competition and Transparency. *Journal of Public Procurement*, 16(4), pp. 528–553. Emerald Publishing.

Hansel (2018a) *Hansel lyhyesti*. Available at: https://www.hansel.fi/tietoa-meista/hansel-lyhyesti/ [Accessed: 25 October 2018].

Hansel (2018b) *Yhteishankinnat*. Available at: https://www.hansel.fi/tietoa-meista/yhteishankinnat/ [Accessed: 29 October 2018].

HE 147/2018 vp *Hallituksen esitys eduskunnalle laiksi Hansel Oy -nimisestä osakeyhtiöstä annetun lain 1 ja 3 §:n muuttamisesta*. Available at: https://www.eduskunta.fi/FI/vaski/KasittelytiedotValtiopaivaasia/Sivut/HE_147+2018.aspx [Accessed: 17 October 2018].

Heinle, M. S., Ross, N., and Saouma, R. E. (2013) A Theory of Participative Budgeting. *The Accounting Review*, 89(3), pp. 1025–1050. American Accounting Association.

Hendrick, R. (1989) Top-down Budgeting, Fiscal Stress and Budgeting Theory. *The American Review of Public Administration*, 19(1), pp. 29–48. SAGE Publications.

Hox, J. J. and Boeije, H. R. (2005) Data Collection, Primary versus Secondary. *Encyclopedia of Social Measurement*, 1, pp. 593–599. Elsevier.

Hyndman, R. J. (2018) *R Documentation for forecast package*. Available at: https://www.rdocumentation.org/packages/forecast/versions/8.4 [Accessed: 4 December 2018].

Hyndman, R. J. and Athanasopoulos, G. (2018) *Forecasting: Principles and Practice*. Available at: https://otexts.org/fpp2/ [Accessed: 13 November 2018].

Ilbeigi, M., Ashuri, B., and Joukar, A. (2017) Time-Series Analysis for Forecasting Asphalt-cement Price. *Journal of Management in Engineering*, 33(1), pp. 1–9. American Society of Civil Engineers.

Karjalainen, K. (2009) Challenges of Purchasing Centralization: Empirical Evidence from Public Procurement. *Doctoral dissertation*. Helsinki School of Economics.

Karjalainen, K., Kivioja, T., and Pellava, S. (2008) Yhteishankintojen kustannusvaikutus: Valtion hankintatoimen kustannussäästöjen selvittäminen. *Helsingin kauppakorkeakoulun julkaisuja.* Helsingin kauppakorkeakoulu.

Knight, L., Harland, C., and Telgen, J. (2012) Public Procurement in Perspective. In *Public Procurement: International Cases and Commentary*. 1st ed. Routledge, pp. 16–24.

Kramer, S. and Hartmann, F. (2014) How Top-down and Bottom-up Budgeting Affect Budget Slack and Performance through Social and Economic Exchange. *Abacus*., 50(3), pp. 314–340. Wiley Online Library.

Kumar, R. R., Patel, A., and Singh, M. (2017) Effect of Procurement Policy on Aid Inflows in the Pacific: Accounting for Economic Growth and Financial Development in Fiji. *Engineering Economics*, 28(5), pp. 505–513.

Lempinen, H. (2013) Design Framework for Performance Management Systems: An Ensemble Approach. *Doctoral dissertation*. Aalto University.

Lloyd, R. E. and McCue, C. P. (2004) What is Public Procurement? Definitional Problems and Implications. In *International Public Procurement Conference Proceedings*. Citeseer, pp. 2–18.

Lütkepohl, H. (2005) *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Matopoulos, A., Bell, J. E., and Aktas, E. (2016) The Use of Modelling in Purchasing/Supply Management Research. *Journal of Purchasing & Supply Management*, 22(4), pp. 262–265. Elsevier.

McCrudden, C. (2004) Using Public Procurement to Achieve Social Outcomes. *Natural Resources Forum*, 28, pp. 257–267.

McCue, C. P. and Pitzer, J. T. (2000) Centralized vs. Decentralized Purchasing: Current Trends in Governmental Procurement Practices. *Journal of Public Budgeting, Accounting & Financial Management*, 12(3), pp. 400–420. Emerald Publishing.

Ministry of Finance (2017) *Valtion hankintakäsikirja 2017*. Available at: https://vm.fi/julkaisu?pubid=20801 [Accessed: 25 October 2018].

Ministry of Finance (2018a) *Julkinen talous*. Available at: https://vm.fi/julkinen-talous [Accessed: 26 October 2018].

Ministry of Finance (2018b) *Valtion hankinnat*. Available at: https://vm.fi/hankinnat [Accessed: 25 October 2018].

Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015) *Introduction to Time Series Analysis and Forecasting*. 2nd ed. John Wiley & Sons.

Oxford Dictionary (2018). Available at: https://en.oxforddictionaries.com/definition/strategic [Accessed: 12 November 2018].

Papadakis, V. M., Lioukas, S., and Chambers, D. (1998) Strategic Decision-making Process: The Role of Management and Context. *Strategic Management Journal*, 19(2), pp. 115–147. John Wiley & Sons.

Reddick, C. G. (2002) Testing Rival Decision–making Theories on Budget Outputs: Theories and Comparative Evidence. *Public Budgeting & Finance*, 22(3), pp. 1–25. Wiley Online Library.

Ruppert, D. and Matteson, D. S. (2015) *Statistics and Data Analysis for Financial Engineering*. 2nd ed. Springer Science & Business Media.

von Rutte, R. (1990) Managing Processes. *The TQM Magazine*, 2(6), pp. 325–329. Emerald Publishing.

Shahandashti, S. M. and Ashuri, B. (2013) Forecasting Engineering News-record Construction Cost Index Using Multivariate Time Series Models. *Journal of Construction Engineering and Management*, 139(9), pp. 1237–1243. American Society of Civil Engineers.

Søreide, T. (2002) Corruption in Public Procurement: Causes, Consequences and Cures. *Research report*. Chr. Michelsen Institute.

Tandberg, D. and Qualls, C. (1994) Time Series Forecasting of Emergency Department Patient Volume, Length of Stay, and Acuity. *Annals of Emergency Medicine*, 23(2), pp. 299–306. Elsevier.

Thai, K. V. (2001) Public Procurement Re-Examined. *Journal of Public Procurement*, 1(1), pp. 9–50. Emerald Publishing.

Thai, K. V. and Grimm, R. (2000) Government Procurement: Past and Current Developments. *Journal of Public Budgeting, Accounting & Financial Management*, 12(2), pp. 231–247. Emerald Publishing.

Trkman, P. (2010) The Critical Success Factors of Business Process Management. *International Journal of Information Management*, 30(2), pp. 125–134. Elsevier.

Tsay, R. S. (2000) Time Series and Forecasting: Brief History and Future Research. *Journal of the American Statistical Association*, 95(450), pp. 638–643. Taylor & Francis.

Wang, Y., Meng, Q., and Tan, Z. (2018) Short-term Liner Shipping Bunker Procurement with Swap Contracts. *Maritime Policy & Management*, 45(2), pp. 211–238. Taylor & Francis.

Witjes, S. and Lozano, R. (2016) Towards a More Circular Economy: Proposing a Framework Linking Sustainable Public Procurement and Sustainable Business Models. *Resources, Conservation, and Recycling*, 112, pp. 37–44. Elsevier.

Appendix A: Interview questions and interviewees

The following interview questions were used to gain an overall description of Hansel's current budgeting process and its deficiencies. Moreover, the interviewees, their roles, and the dates of the interviews are listed in Table A.1.

1. What is your role in the current budgeting process?
2. What do you think are the strengths and weaknesses of the current process?
3. How would you improve the current process?
4. Do you feel the current process gives adequate conditions for making decisions about the framework agreement volumes, and based on them the service fees?
5. What kind of tools would you desire to help make decisions about the framework agreement volumes?
6. If the prediction of framework agreement volumes is enhanced with forecasting models, at which point of the process you think the models would be used and by whom?

Table A.1 The interviewees and the dates of the interviews

Interviewee	**Role**	**Date**
Halonen Mikko	Controller	21.11.2018
Hietaranta Kalle	Head of Unit in Category Management	28.11.2018
Jokela Heli	Controller	23.11.2018
Närvänen Susanna	Chief Category Officer	7.12.2018
Olkinuora Mervi	Head of Unit in Category Management	28.11.2018
Pursimo Juho	Analyst	26.11.2018

Appendix B: Mathematical properties of ARIMA models

Both AR(p) and MA(q) models can be expressed as the first-order system in equations (B.1) and (B.2) and further in a matrix form in equations (B.3) and (B.4). Especially the matrix forms ease modelling and forecasting as they simplifie complex lag structures to resemble AR(1) and MA(1) models.

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix} = \begin{pmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{B.1}$$

$$\begin{pmatrix} y_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \delta \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q+1} \end{pmatrix} - \begin{pmatrix} \theta_1 & \theta_2 & \dots & \theta_q \\ 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \varepsilon_{t-q} \end{pmatrix} \tag{B.2}$$

$$Y_t = \Delta + \Phi Y_{t-1} + E_t \tag{B.3}$$

$$Y_t = \Delta + E_t - \Theta E_{t-1} \tag{B.4}$$

Stationarity and invertibility

Weak stationarity is mathematically defined as in equations (B.5) - (B.7) through the mean, variance, and autocovariance, respectively (see e.g. Box *et al.*, 2008).

$$E[y_t] = E[y_{t-j}] = \mu \tag{B.5}$$

$$Var[y_t] = Var[y_{t-j}] = \sigma^2 \tag{B.6}$$

$$Cov[y_t, y_{t-i}] = Cov[y_{t-j}, y_{t-j-i}] = \gamma_i \tag{B.7}$$

Consequently, the mean and the variance are constants and the autocovariance between two observations only depends on the lag *i*. In practice, the mean, variance, and autocovariance are estimated by the sample mean $\tilde{y}$, the sample variance $\hat{\sigma}_y^{\ 2}$, and the sample autocovariance $\hat{\gamma}_i$ (see e.g. Montgomery *et al.*, 2015).

For a stationary AR(1) the unconditional moments are (see e.g. Montgomery *et al.*, 2015; Ruppert and Matteson, 2015)

$$E[y_t] = \frac{\delta}{1-\phi} \tag{B.8}$$

$$Var[y_t] = \frac{{\sigma_\varepsilon}^2}{1-\phi^2} \tag{B.9}$$

$$Cov[y_t, y_{t-i}] = \phi^i Var[y_t]. \tag{B.10}$$

For a stationary AR(p) the unconditional moments can be depicted similarly to equations (B.8) - (B.10) with the matrix form in equation (B.3).

The simplest example of a stationary process is white noise (Ruppert and Matteson, 2015). Since the white noise error term of AR models is a sequence of uncorrelated random variables and it has a zero mean, it is stationary. Furthermore, if the variables are independent and identically distributed (i.i.d.), the white noise process is strictly stationary (Box *et al.*, 2008). In addition, if the variables are independently normally distributed, the process is Gaussian white noise (Montgomery *et al.*, 2015). As a stationary process white noise has the unconditional moments as in equations (B.11) - (B.13) (see e.g. Ruppert and Matteson, 2015).

$$E[y_t] = 0 \tag{B.11}$$

$$Var[y_t] = \sigma^2 \tag{B.12}$$

$$Cov[y_t, y_{t-i}] = 0, \text{for } i \neq 0 \tag{B.13}$$

The stationarity and invertibility of an ARIMA(p,d,q) corresponds to the stationarity of an AR(p) and the invertibility of an MA(q). An ARIMA(p,d,q) is stationary and invertible when the eigenvalues of both Φ and Θ in equation (6) lie inside the unit circle, i.e. equations (B.14) and (B.15) hold (see e.g. Box *et al.*, 2008).

$$|\lambda_k| < 1 \text{ for all } k = 1, \ldots, r \text{ where } r = \max(p,q) \tag{B.14}$$

$$|\zeta_k| < 1 \text{ for all } k = 1, \ldots, r \text{ where } r = \max(p,q) \tag{B.15}$$

Ljung-Box test

The Ljung-Box test is a statistical test for jointly testing if the *i*-order sample autocorrelations are zero and a process is white noise. The test statistic is presented in equation (B.16) (see e.g. Montgomery *et al.*, 2015). The null hypothesis of the sample autocorrelations being jointly zero is rejected if the Q statistic is larger than the $\chi^2(k)$ critical value or the p-value is smaller than the desired significance level. In the equation T equals to the sample size and k is the number of lags being tested.

$$Q = T(T+2)\sum_{i=1}^{k} \frac{Corr[\hat{y_t}, y_{t-i}]^2}{T-i} \sim \chi^2(k) \tag{B.16}$$

Trends and seasonality

A time series which exhibits a trend can be divided into a trend component τ_t and a cyclical component c_t, as shown in equation (B.17). The trend component can be removed by fitting to the data a regression model which describes the trend component as in equation (B.18) and estimating the stationary cyclical component from the regression residuals $\hat{c}_t$ in equation (B.19) (see e.g. Montgomery *et al.*, 2015).

$$y_t = \tau_t + c_t \tag{B.17}$$

$$y_t = (\alpha + \beta t) + c_t \tag{B.18}$$

$$\hat{c}_t = y_t - (\hat{\alpha} + \hat{\beta}\, t) \tag{B.19}$$

When the trend component is not assumed to be deterministic, the cyclical component in equation (B.17) is related to the first difference of y_t. To remove the trend component, a new time series is obtained by applying a difference operator ∇ to the original time series. Hence, the first and the second difference correspond to equations (B.20) and (B.21), respectively (see e.g. Montgomery *et al.*, 2015; Ruppert and Matteson, 2015).

$$\nabla y_t = y_t - y_{t-1} \tag{B.20}$$

$$\nabla^2 y_t = \nabla y_t - \nabla y_{t-1} \tag{B.21}$$

A seasonal component is removed from a time series with seasonal differencing. The lag-*s* seasonal difference operator ∇_s in equation (B.22) captures the difference between the current value and the *s* lags before value of the variable of interest (see e.g. Box *et al.*, 2008; Ruppert and Matteson, 2015). The value of *s* is selected according to the type of the time series, e.g. $s = 4$ for quarterly data.

$$\nabla_s y_t = y_t - y_{t-s} \tag{B.22}$$

Parameter estimation

Maximum likelihood and conditional least squares can be used to estimate both linear and nonlinear models. Maximum likelihood estimates the model parameters by maximising the likelihood function, i.e. the joint probability density function of the sample (see e.g. Ruppert and Matteson, 2015). Often the parameters are estimated with the logarithm of the likelihood function, since it makes computations easier and gives the same estimators as when using the likelihood function (Montgomery *et al.*, 2015).

Conditional least squares is a somewhat simpler estimator than maximum likelihood as it maximises the logarithm of the conditional density function, given in equations (B.23) and (B.24) (see e.g. Ruppert and Matteson, 2015). In both equations *T* equals to the sample size and *p* to the number of parameters in the AR part of the model. Neither of the methods have an analytical solution, and their numerical solution is computed by iterating the value of log-likelihood until no greater value is obtained.

$$L = (\frac{1}{\sqrt{2\pi}\sigma})^{T-p} \prod_{t=p+1}^{T} \exp(-\frac{\varepsilon_t^2}{2\sigma^2}) \tag{B.23}$$

$$\log L = -\frac{T-p}{2}\log 2\pi - \frac{T-p}{2}\log \sigma^2 - \frac{1}{2}\sum_{t=p+1}^{T}\frac{\varepsilon_t^2}{\sigma^2} \tag{B.24}$$

Model selection

The likelihood ratio (LR) test in equation (B.25) is used to compare the goodness of fit of two or more models and it is based upon the log-likelihood function in equation (B.24). The test compares the log-likelihoods $\log \hat{L}_0$ and $\log \hat{L}_1$ of a restricted and an unrestricted model, respectively, to determine the correct lag structure. (Ruppert and Matteson, 2015)

$$LR = 2(\log \hat{L}_1 - \log \hat{L}_0) \sim \chi^2(k) \tag{B.25}$$

The Akaike's (AIC) and Bayesin (BIC) information criteria in equations (B.26) and (B.27) use also the log-likelihood function in equation (B.24) to evaluate the fit of different models. In equations (B.26) and (B.27) p and q equal to the number of parameters in the ARMA(p,q) model, and T to the sample size. Instead of being statistical tests, the information criteria are numerical criteria reflecting the trade-off between the fit and the parsimony of a model. The terms $2(p+q)$ in equation (B.26) and $\log T\,(p+q)$ in equation (B.27) penalise models with many lags and are thus called complexity penalties. Further, BIC selects typically smaller number of lags since it penalises model complexity more than AIC as $\log T > 2$ when $T > 8$. (Ruppert and Matteson, 2015)

$$\mathrm{AIC} = -2\log L + 2(p+q) \tag{B.26}$$

$$\mathrm{BIC} = -2\log L + \log T\,(p+q) \tag{B.27}$$

Forecasting and forecast accuracy

The core concepts of forecasting with ARIMA models are the *i*-step-ahead forecast, the *i*-step-ahead forecast error, and the variance of the forecast error. These are calculated for an ARIMA(p,d,q) process such as in equations (B.28) - (B.30) (see Montgomery *et al.*, 2015).

$$\hat{y}_{t+i} = E[y_{t+i}|y_t, y_{t-1}, \ldots] = \delta + \sum_{j=i}^{\infty} -\theta_j \varepsilon_{t+i-j} \tag{B.28}$$

$$e_t = y_{t+i} - \hat{y}_{t+i} = \sum_{j=0}^{i-1} - \theta_j \varepsilon_{t+i-j} \tag{B.29}$$

$$Var[e_t] = \sigma^2 \sum_{j=0}^{i-1} - \theta^2 \tag{B.30}$$

The mean forecast error (ME), the mean absolute forecast error (MAE), and the mean squared forecast error (MSE) in equations (B.31) - (B.33) are generally used to specify the accuracy of a forecast. To gain measures of forecast accuracy that are independent of the unit of the original series, the relative forecast error (RE) in equation (B.34) is used. (see Montgomery *et al.*, 2015)

$$ME = \frac{1}{T} \sum_{t=1}^{T} e_t \tag{B.31}$$

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |\, e_t| \tag{B.32}$$

$$MSE = \frac{1}{T} \sum_{t=1}^{T} {e_t}^2 \tag{B.33}$$

$$RE = (\frac{e_t}{y_t}) * 100 \tag{B.34}$$

Appendix C: Framework agreement subtotals

Table C.1. Total volumes of the framework agreement subtotals in 2017

Framework agreement subtotal	Total volume (EUR)
Occupational Health Care Services	61 195 913,6
Electricity	60 481 199,6
Computers	53 658 031,6
IT Consulting	51 938 885,1
Facility Services for Premises	47 897 190,0
Data Centre and Capacity Services	40 571 646,5
Leasing Services	40 054 770,5
Cars	37 740 537,8
Microsoft	34 961 044,6
Scheduled Flights	31 635 638,4
Office Furniture	29 306 001,5
Fuel	28 417 844,8
National Accommodation and Conference Services	27 849 414,0
Teleoperator Services	20 418 796,3
Management Consulting	16 539 046,1
Office and IT Supplies	14 354 621,3
Vehicle Leasing Services	13 058 484,4
Groceries and Non-food Products	12 863 701,8
Rail Transport Services	12 101 912,3
Human Resources Management Services	11 694 954,5
Cell Phones	10 991 517,1
IT Network Equipment	10 679 539,0
Printing Services	10 215 322,3
Security Technology and Services	10 131 988,0
Projection Techniques	10 002 158,3
Periodical Publications’ Supply Services	9 701 020,7
Printing	9 606 994,8
Tools and Supplies	8 040 727,9
Fuel Purchases from Distribution Substations	7 762 277,9
Car Rental	7 030 507,4
Maintenance and Information Services for Cars	6 385 348,9
Charter Bus Services	6 327 678,7
Data Storage Solutions	5 058 501,0
Servers	5 055 048,1
Machinery	4 480 145,5
Interpretation Services	4 472 206,8
Communication and Marketing Services	4 396 071,0
Video Conference Services	4 263 743,2
Data Communication Services	3 974 931,8
Electricity and HVAC Equipment	3 862 057,1
Translation Services	3 736 031,9
Auditing	3 612 711,0
Gases	2 866 822,4
Maintenance Flights	2 686 461,6
Travel Agency Services	2 539 640,8
Charter Flights	2 406 669,1
Removal Transport Services	2 088 905,4
Agency Service of Foreign Literature	1 742 017,5
Agency Service of Finnish Literature	1 720 249,3

Table C.1. (continued)

International Accommodation Services	1 559 547,9
Heavy Motor Vehicles	1 481 034,4
Low-powered Motor Vehicles	1 308 516,7
ICT Education Services	1 271 966,4
Red Hat and Suse Distribution Channel	1 258 101,3
Adobe	1 073 832,7
Point of Sale System	973 130,6
Travel Insurances	941 148,3
Cruise Services	894 596,5
Safes and Fireproof Safes	770 511,1
Information Security Equipment	617 424,5
Translation Services for State Council	552 296,0
Vehicle Insurances	494 440,0
e-Tendering	361 252,0
Locking Systems and Services	218 427,5
Uninterruptible Power Supplies (UPS)	122 523,9
License Management Services	6 232,0

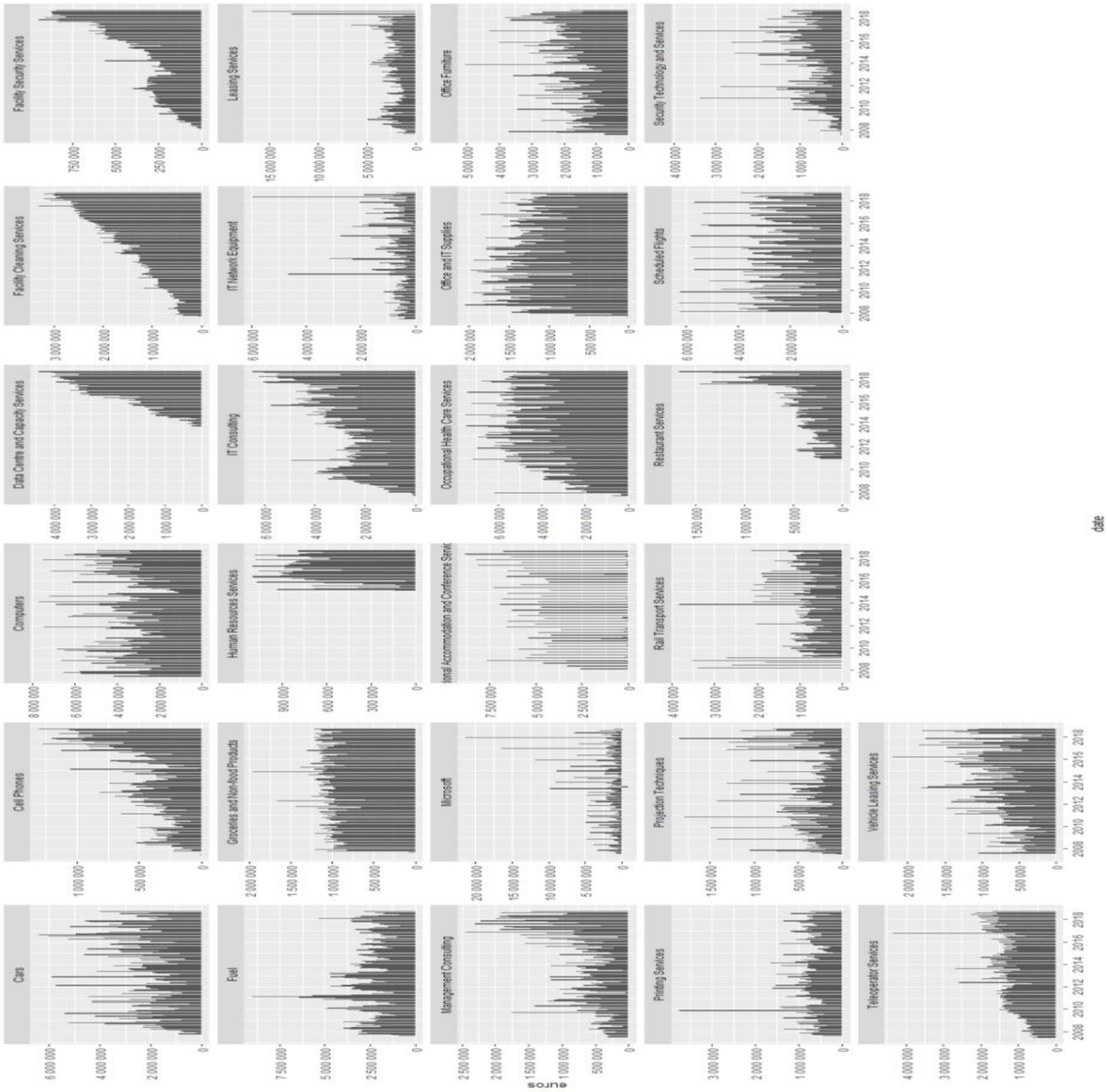
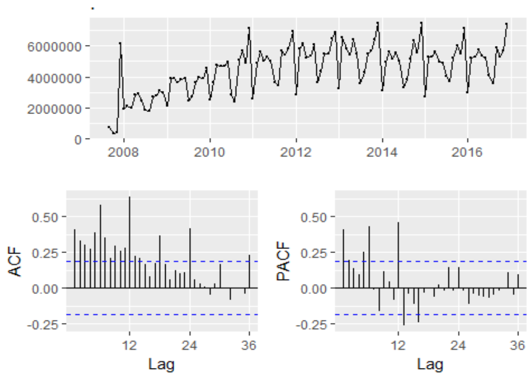


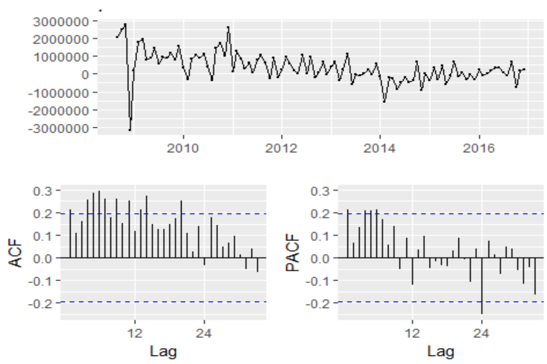
Figure C.1. Plots of the modelled framework agreement subtotals

Appendix D: Example code of modelling

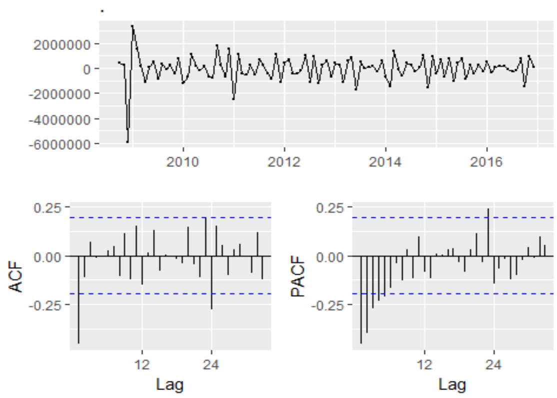
```
healthcare <- read_rds("data/health care.rds") %>%
  transmute(date = rapo_pvm, euros = eur)

hc_in <- healthcare %>% filter(!(str_detect(date, "2017") | str_detect(date, "2018")))
%>%  as.data.frame()
hc_in <- ts(hc_in[1:112,2], frequency = 12, start = c(2007, 9))

hc_out <- healthcare %>% filter(str_detect(date, "2017") | str_detect(date, "2018"))
%>%  as.data.frame()
hc_out <- ts(hc_out[1:22,2], frequency = 12, start = c(2017,1))
hc_b <- read_rds("data/budjetit/health care.rds")

hc_in %>% ggtsdisplay()
```




```
hc_in %>% diff(lag=12) %>% ggtsdisplay()
```

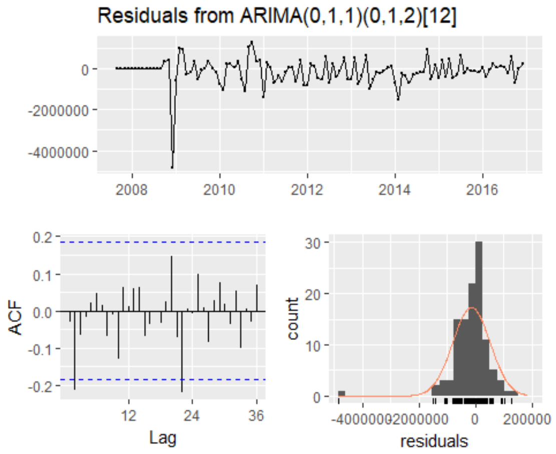



```
hc_in %>% diff(lag=12) %>% diff() %>% ggtsdisplay()
```



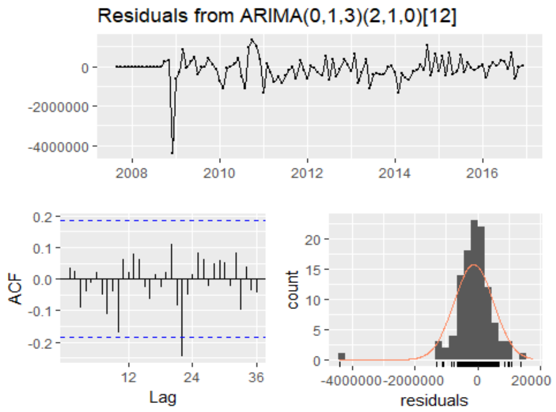
```
hc_m1 <- auto.arima(hc_in)
hc_m2 <- auto.arima(hc_in, stepwise = FALSE, approximation = FALSE)
hc_m3 <- hc_in %>% Arima(order = c(2,1,1), seasonal = c(0,1,2))
hc_m4 <- hc_in %>% Arima(order = c(2,1,0), seasonal = c(0,1,2))
hc_m5 <- hc_in %>% Arima(order = c(5,1,1), seasonal = c(0,1,2))
hc_m6 <- hc_in %>% Arima(order = c(5,1,0), seasonal = c(0,1,2))
```

```
checkresiduals(hc_m1)
```



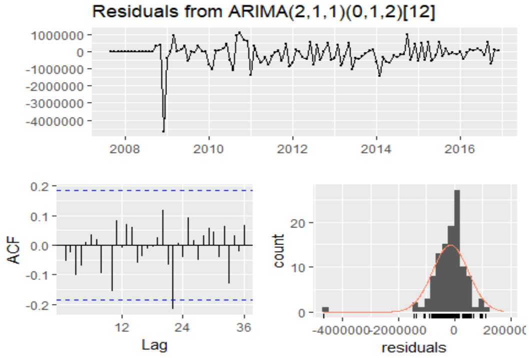

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(0,1,2)[12]
## Q* = 21.931, df = 21, p-value = 0.4035
##
## Model df: 3.   Total lags used: 24
```

```
checkresiduals(hc_m2)
```



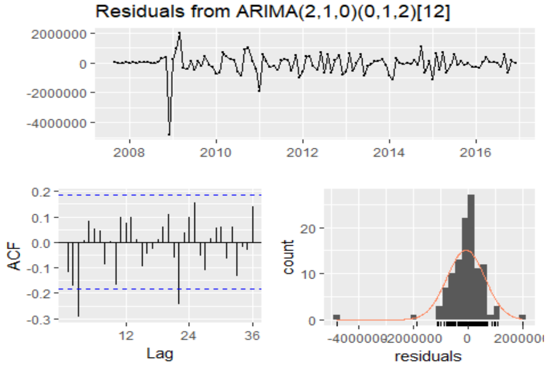

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,3)(2,1,0)[12]
## Q* = 21.469, df = 19, p-value = 0.3115
##
## Model df: 5.   Total lags used: 24
```

checkresiduals(hc_m3)



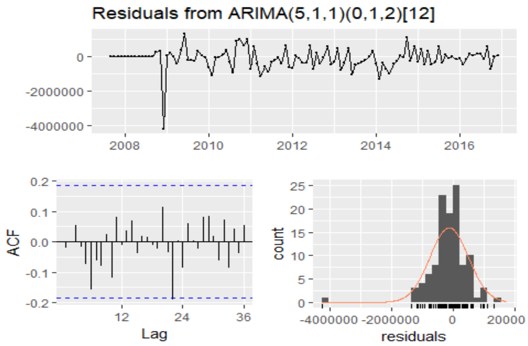
##
Ljung-Box test
##
data: Residuals from ARIMA(2,1,1)(0,1,2)[12]
Q* = 18.835, df = 19, p-value = 0.4675
##
Model df: 5. Total lags used: 24

checkresiduals(hc_m4) #not white noise -> model not a good fit



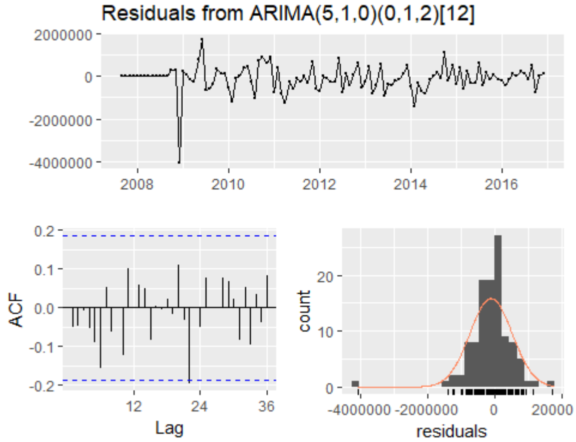
##
Ljung-Box test
##
data: Residuals from ARIMA(2,1,0)(0,1,2)[12]
Q* = 38.794, df = 20, p-value = 0.007072
##
Model df: 4. Total lags used: 24

checkresiduals(hc_m5)



```
##
## 	Ljung-Box test
##
## data:  Residuals from ARIMA(5,1,1)(0,1,2)[12]
## Q* = 17.501, df = 16, p-value = 0.3539
##
## Model df: 8.   Total lags used: 24
```

checkresiduals(hc_m6)




```
##
## 	Ljung-Box test
##
## data:  Residuals from ARIMA(5,1,0)(0,1,2)[12]
## Q* = 18.2, df = 17, p-value = 0.3764
##
## Model df: 7.   Total lags used: 24
```

hc_f1 <- **forecast**(hc_m1, h = 22)
hc_f2 <- **forecast**(hc_m2, h = 22)
hc_f3 <- **forecast**(hc_m3, h = 22)
hc_f5 <- **forecast**(hc_m5, h = 22)
hc_f6 <- **forecast**(hc_m6, h = 22)

accuracy(hc_f1, hc_out)

```
##                     ME     RMSE      MAE      MPE      MAPE      MASE
## Training set -149104.5 672366.1 407955.9 -4.86008 10.049754 0.6281496
## Test set     -192927.0 350666.7 264815.6 -3.96771  5.293879 0.4077494
##                     ACF1 Theil's U
## Training set -0.03071365        NA
## Test set     -0.47425237 0.2390433
```

accuracy(hc_f2, hc_out)

```
##                      ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -124619.39 640137.0 394849.7 -4.183540 9.551394 0.6079692
## Test set       71255.32 328269.5 271828.8  1.649115 5.534167 0.4185479
##                     ACF1 Theil's U
## Training set  0.03356772        NA
## Test set     -0.48399540 0.2454049
```

```
accuracy(hc_f3, hc_out)
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -148929.5 656502.0 397774.0 -4.788815 9.652273 0.6124720
## Test set     -180830.9 347126.3 263129.7 -3.734006 5.248232 0.4051535
##                    ACF1 Theil's U
## Training set -0.05368959        NA
## Test set     -0.45369494 0.2376901

accuracy(hc_f5, hc_out)
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -126250.2 633280.2 401262.0 -3.975450 9.342205 0.6178426
## Test set     -189551.4 347256.4 256140.9 -4.063987 5.195717 0.3943926
##                    ACF1 Theil's U
## Training set -0.02162842        NA
## Test set     -0.47604691 0.2336279

accuracy(hc_f6, hc_out)
##                    ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -107486.5 636129.5 409616.9 -3.635622 9.700833 0.6307069
## Test set     -229117.5 377097.6 291762.9 -4.834789 5.871816 0.4492415
##                    ACF1 Theil's U
## Training set -0.05038425        NA
## Test set     -0.43062992 0.2535443

accuracy(as.numeric(hc_b), hc_out)
##                ME     RMSE      MAE      MPE     MAPE       ACF1 Theil's U
## Test set 95491.39 413738.9 344608.6 1.099437 7.113493 -0.2068176 0.3118777

hc_2 <- hc_f2$mean %>% as_tibble() %>% transmute(f2 = x)
y <- data.frame(n = 1:22)
hc_f <- bind_cols(y, hc_2, as_tibble(hc_out), as_tibble(as.numeric(hc_b)))

ggplot(hc_f, aes(n, y = value, colour = forecasts)) +
  geom_line(aes(y = f2, col = "f2")) +
  geom_line(aes(y = x, col = "actual")) +
  geom_line(aes(y = value, col = "budgeted"))
```