**JUDGMENTAL RECONCILIATION**

**OF HIERARCHICAL FORECASTS**

Violina Sarma

Cardiff University

This dissertation is submitted to Cardiff University

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

September 2015

|  |  |
| --- | --- |
| **Candidate’s ID Number** | **1458141** |
| **Candidate’s Title** | **Please circle appropriate value**  **Mr / Miss / Ms / Mrs / Rev / Dr / Other please specify …………………..** |
| **Candidate’s Surname** | **Sarma** |
| **Candidate’s First Names** | **Violina** |

**DECLARATION**

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed …………………………………………………………. (candidate) Date …………………………

**STATEMENT 1**

This dissertation is being submitted in partial fulfilment of the requirements for the degree of …………………………(insert MA, MSc, MBA, MScD, LLM etc, as appropriate)

Signed …………………………………………………………. (candidate) Date …………………………

**STATEMENT 2**

This dissertation is the result of my own independent work/investigation, except where otherwise stated.

Other sources are explicitly acknowledged with references. A Bibliography is appended.

Signed …………………………………………………………. (candidate) Date …………………………

**STATEMENT 3 – To be completed where the second copy of the dissertation is submitted in an approved electronic format**

I confirm that the electronic copy is identical to the bound copy of the dissertation

Signed …………………………………………………………. (candidate) Date …………………………

**STATEMENT 4**

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …………………………………………………………. (candidate) Date …………………………

**STATEMENT 5 - BAR ON ACCESS APPROVED**

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access approved by the Graduate Development Committee.**

Signed …………………………………………………………. (candidate) Date …………………………

**Abstract**

Statistical forecasts in business are often subject to judgmental interventions. Managers make judgmental adjustments at different levels of a hierarchy. As a result, these adjusted forecasts are not ‘aggregate consistent’ across the hierarchy. The aim of this thesis is to explore how these forecasts can be reconciled through an interpersonal reconciliation process among different managers. It explores the different factors that might affect this judgmental reconciliation process, as well as comparing the accuracy of this approach against standard statistical aggregation methods.

The sales data of three products (*iPhone*, *iPad* and *Mac*) of the *Apple* company are organised in two-level hierarchical structures based on product-type. A laboratory experiment is conducted where the subjects are provided with the historical sales data of each product, quarterly statistical forecasts for 2015 and contextual information regarding expected sales of these products over 2015. They first judgmentally adjust the given statistical forecasts individually and then reconcile these adjusted forecasts at a group level. From the group discussions, the main factors affecting the reconciliation process are identified. Deviations from individual to group round forecasts are calculated to investigate whether these deviations are related to forecast accuracy. Finally, forecast accuracy evaluations are performed to determine the best method for aggregating judgmentally adjusted statistical forecasts.

**Keywords:** Hierarchical forecasting, Cross-sectional aggregation, Judgmental reconciliation

**Acknowledgements**

I would like to use this acknowledgement to express my profound gratitude to everyone to whom the rest of the dissertation owes its existence. First, I would like to offer my sincere thanks to my supervisors Prof. Aris A. Syntetos and Dr. Fotios Petropoulos who have been a fountainhead of relentless encouragement for me throughout the dissertation. Their meticulous planning, constructive criticism, and enlightening guidance on the exciting area of mixed method research strategies kept me motivated and helped me sail through the ebbs and flows of the journey so far.

My heartfelt gratitude is also directed towards Prof. Mike Wallace for his overwhelming support throughout the academic year. His sustained efforts and patience in assisting me with the qualitative strand of the research work provided me with the guiding light that exalted this dissertation to the next level. I am deeply thankful to the participants (Masters’ students and young researchers from LOM section) of the experiment whose contributions to the research on judgmental forecasting were instrumental in this dissertation. I would also like to thank all my LOM colleagues with whom I had the pleasure of sharing this wonderful and enriching learning experience.

Finally, I would like to acknowledge the funding from the President Scholarship of Cardiff University without which none of this research work would have been possible.

Table of Contents

**Declaration and Statementi**

**Abstractii**

**Acknowledgementsiii**

**List of appendicesvi**

**List of tablesvi**

**List of figuresvii**

**Chapter 1: Introduction1**

1.1 Research background1

1.2 Research aim and research questions 3

1.3 Structure of the thesis 5

**Chapter 2: Critical review of the literature7**

2.1 Judgmental forecasting7

2.1.1 Unaided and aided judgmental forecasting9

2.1.2 Empirical and experimental research10

2.1.3 Group judgmental forecasting12

2.2 Reconciliation of hierarchical forecasts or cross-sectional aggregation12

2.2.1 Bottom-Up15

2.2.2 Top-Down15

2.2.3 Middle-Out16

2.2.4 Optimal reconciliation16

2.3 Research gaps18

**Chapter 3: Research Methodology21**

3.1 Epistemological philosophy21

3.2 Research design25

3.2.1 Experimental design25

3.2.2 Research sample29

3.2.3 Data collection methods30

3.2.4 Ethics33

3.2.5 Pilot study34

3.3 Accuracy evaluation34

**Chapter 4: Analysed findings 38**

4.1 Qualitative data analysis38

4.2 Quantitative data analysis43

4.2.1 Individual versus group forecasts43

4.2.2 Judgmental reconciliation versus statistical reconciliation 46

4.2.3 Derivation and forecast performance50

**Chapter 5: Discussion 53**

5.1 Research question 153

5.2 Research question 255

**Chapter 6: Summary and conclusion 58**

6.1 Summary58

6.2 Implications for theorists and practitioners of forecasting58

6.3 Research contributions59

6.4 Limitations and future directions60

**Bibliography 62**

**Appendices 68**

**List of appendices**

Appendix A: R code for statistical forecasts for questionnaires68

Appendix B: R code for statistical aggregation methods and error measures70

Appendix C: R code for calculation of AvgRelMAE77

Appendix D: R code for deviation evaluation79

Appendix E: R code for comparison between deviation and forecast performance80

Appendix F: Approved ethics form81

Appendix G: Questionnaires93

**List of tables**

Table 3.1: Paradigm contrast table24

Table 3.2: Description of statistical forecasting methods27

Table 3.3: Dimensions of contrast between three methodological positions33

Table 4.1: Accuracy measurement of different reconciliation methods for the four nodes48

Table 4.2: Spearman’s rank correlation coefficient between deviation and D\_APE for different dimensions52

**List of figures**

Figure 1.1: Forecasting steps3

Figure 1.2: Pictorial description of the thesis structure5

Figure 2.1: Hierarchy of a clothing company13

Figure 2.2: Hierarchical attribute of demand data14

Figure 2.3: A hierarchical structure with three levels17

Figure 2.4: A two-level hierarchy based on geographical locations20

Figure 3.1: Inductive-Deductive research cycle23

Figure 3.2: A two-level business hierarchy26

Figure 3.3: Historical sales of Apple, iPhone, iPad and Mac28

Figure 3.4: Graphic illustration of parallel mixed designs32

Figure 4.1: Representation of the four themes39

Figure 4.2(a): Boxplot of percentage deviations categorised by product44

Figure 4.2(b): Boxplot of percentage deviations categorised by group45

Figure 4.3: Histograms of individual forecasts, group forecasts and deviations 46

Figure 4.4: MAPE across the groups for different reconciliation methods for quarters 1, 2 and both 47

Figure 4.5: Average Relative MAE of group reconciliation method relative to the other methods49

**Chapter 1: Introduction**

* 1. **Research background**

Forecasting is an attempt to estimate the future by examining the past patterns. It is the development and application of different models to extrapolate the past-observed data into the future (Winkler 1987). Business forecasting forms an integral part of different types of management decision making in organisations. It is important for an organisation to make timely decisions in the face of uncertainty (Hanke and Reitsch 1998, p. 507). The dynamic and uncertain nature of the business world makes forecasting the springboard for all major management decisions (Gross and Peterson 1983, p. 392). The more forecasting is done, lower is the future uncertainty (Makridakis and Wheelwright 1987, p. 11). Organisations need to estimate the future demand of their products to plan their operations (Ord and Filed 2013, p. 6). A good forecasting procedure is necessary as planning involves making decisions that have effects on the future (Firth 1977, p. 1). When a forecaster makes a decision about the forecast, a series of variables are to be taken into consideration like statistical characteristics of the data, the forecast lead time, the aggregation level (Fildes 1987, p. 162). When a stable relationship is found among these variables, then the focus shifts to the forecast accuracy. However, errors are an intrinsic part of any forecasting procedure (Hanke and Reitsch 1998, p. 1). The predictions can rarely be the same as the future outcome, so the aim of the forecasters is to minimise this inevitable error as much as possible. A forecast in itself is of no value to an organisation and needs to be converted into valuable information. It helps in establishing useful goals for future and determining how these goals are to be achieved (Gross and Peterson 1983, p. 2). Therefore, forecasts serve as a guide to the planning process.

There are a number of statistical methods that generate forecasts based on historical data. A number of software packages are available in the market that help produce these quantitative or statistical forecasts (SFs). Due to a number of factors like shorter product life cycles, world globalisation, competitive and aggressive marketing; sales forecasting has become more complex (Trapero et al. 2012). The past patterns are not always good indicators of the future as they can overestimate or underestimate the future (Gross and Peterson 1983, p. 31). In addition, managers possess expert knowledge and inside information (like future promotions) that are not incorporated by the statistical methods. To add this information to the forecast, managers make forecasts using their own judgments. This method of forecasting using subjective opinions of experts is referred to as ‘judgmental forecasting’. It is defined as “the process of producing forecasts based on combining information and the subjective beliefs of individuals” (Ord and Fildes 2013, p.361). This is also called the qualitative aspect or method of forecasting. The use of judgment is an essential component for all forms of forecasting (Hanke and Reitsch 1998, pp. 494-495). There are cases when judgmental forecasting is the only forecasting option for instance when there is no historical data available, the data is incomplete or is delayed, new products are being launched, unique market conditions or new competition in the market (Hyndman and Athanasopoulos 2013a). This method can be applied either informally or in a structured manner. It has been found that a combination of both statistical methods and human judgment potentially improves the forecast accuracy, rather than any one of them in isolation (Blattberg and Hoch 1990; Hanke and Reitsch 1998, p. 1; Lawrence et al. 2006; Trapero et al. 2012). Judgmental adjustments to the SFs are very common in practice today. This process of human intervention into the forecasting process and making adjustments to SFs is outlined in figure 1.1. The historical data is used to produce the quantitative forecasts by the statistical tools. On the other hand, experts use the non-historical contextual data to make adjustments to the SFs based on judgments. This results in the adjusted final forecasts that are used for the planning purposes. Although it is not shown in figure 1.1, the decision makers may indeed rely upon historical data (not necessarily quantitative information that is reflected on the time-series) when making adjustments. As business forecasting is never devoid of management judgment, it is considered “*an art woven into science and principles teamed with pragmatism*” (Ord and Fildes 2013, p. xvii).

One of the most important challenges in modern supply chain management is that of demand uncertainty (Syntetos et al. 2015a). In response, organisations emphasise on making improved decisions at different levels of aggregation. Different strategic and operational divisions of decision making processes are sub-divided for ease of handling according to aggregation levels like location, product and time. This implies the need of demand forecasting at different aggregation levels. For example, decisions concerned with inventory control require demand forecasts per product and location. There can be one department that does the aggregate planning for the entire company and a department that does planning for individual products. Time-series data can be disaggregated based on a number of characteristics like geographical location, product-type or SKU. This disaggregated data structure forms a hierarchy, which enables logical methods of aggregation. Cross-sectional aggregation adds up the forecasts across different products or locations in a hierarchy for the same time periods. The significance of this method is it utilizes the hierarchical characteristics of the data (Syntetos et al. 2015a). The different statistical methods used for cross-sectional aggregation are the bottom-up, top-down, middle-out and optimal reconciliation methods. However, subjective judgment is needed to make reflections of reality in forecasting. Managers make judgmental adjustments to incorporate all relevant available information (Winkler 1987, p. 248). However, when statistical reconciliation methods are used, the adjustments can be performed only at one aggregate level, either at the most aggregated or the most disaggregated level (Syntetos et al. 2015a). These methods do not allow judgmental interventions into all levels of hierarchical forecasting.

DATA

History data

Forecasting TOOL

FORECASTER

Adjustment review

Adjusted Forecast

Non-history data

Figure 1.1: Forecasting steps (adapted from Lawrence et al. 2006)

**1.2 Research aim and research questions**

Judgmental reconciliation of the adjusted forecasts can be one option of hierarchical aggregation. It can be achieved through group decision making while attaining consensus regarding the reconciled forecasts. From a theoretical perspective, many researchers have studied the statistical reconciliation methods but there is not much research on the group reconciliation techniques (Syntetos et al. 2015a). From a practical perspective, managers belonging to different hierarchical levels are unlikely to agree on one single aggregation method. The dynamics of the real world can be best captured in inter-personal reconciliation process. Many researchers have called for future researches in this field of judgmental reconciliation process (Spithourakis et al. 2015, Syntetos et al. 2015a). This judgmental reconciliation process has a number of advantages over the statistical methods. It allows incorporating a range of managerial viewpoints into the discussion (Ord and Fildes 2013, p. 367). Judgment or opinions can be injected at all levels of hierarchy. The forecasters manually fix the differences in the forecasts of different levels. Through this method, they attain consensus and a sense of “ownership” of the forecasts in a collective manner (Petropoulos 2014). In this process of group decision making, the chances of overlooking an event and course of action is diminished (Goodwin and Wright 2014, p. 390). Additionally, it is a good practice to adjust the independent individual forecasts prior to group interaction (Sniezek 1990). Along with the advantages, the group consensus method also has a number of drawbacks. The group dynamics can distort the process of attaining consensus. In hierarchical organisations, the high-ranking managers can dominate or influence the group decisions (Gross and Peterson 1983, p. 57). The weight assigned to each individual’s assessment will depend on the role and personality of that manager in the organisation (Makridakis and Wheelwright 1989, p. 241). Another drawback is individuals usually favour their own judgment over that of the others. Moreover, the group members are more likely to make risky decision collectively than individually, as the responsibility will be “shared” (Makridakis and Wheelwright 1989, p. 252). Therefore, it is necessary to understand the dynamics of interactions in cohesive groups to avoid the drawbacks and problems of group forecasting (Makridakis and Wheelwright 1989, p. 250). These arguments demonstrate the importance of this research from an academic and practitioner perspective.

The aim of this research is to see how consensus is achieved through group interactions. The objective is to understand the reconciliation process and identify the different elements of group decision making process. With the motivation that “*the synergy of individuals may make the overall quality of the group decision greater than the sum of the parts*” (Goodwin and Wright 2014, p.390), the following research questions are framed:

**RQ1.** *What factors affect the process of judgmentally reconciling hierarchical forecasts?*

**RQ2.** *Does judgmental reconciliation result in better forecast accuracy than statistical reconciliation methods?*

With an attempt to address the above research questions, an experiment is designed to see how group consensus is achieved while reconciling hierarchical forecasts judgmentally. Participants are divided into groups of four and each representing a node of a 1-3 pyramid business hierarchy. The participants use their forecasting knowledge to make judgmental adjustments to SFs individually based on the contextual information provided. Then they reconcile these individual forecasts with the other group members through consensus. The group discussions are observed and audio-recorded to determine how the participants move from individual to group forecasts. From these discussions, the main factors affecting the judgmental reconciliation process are identified. The forecast accuracy of the judgmentally reconciled forecasts are compared to those of statistical reconciled forecasts from different statistical reconciliation methods. This further leads to the recommendation of the best reconciliation method.

**1.3 Structure of the thesis**

The structure of this thesis is constructed as described in figure 1.2.

Introduction

Literature review

Methodology

Analysed findings

Discussion

Summary and Conclusion

Figure 1.2: Pictorial description of the thesis structure

This first chapter provided an overview of the research background to judgmental reconciliation of hierarchical forecasts. The motivation behind this study and the research questions were outlined in detail. The next chapter reviews the past literature on judgmental forecasting, focusing chiefly on cross-sectional aggregation of forecasts. Both primary and secondary data collection methods are used to address the two research questions under study. Chapter 3 presents the methodology employed for data collection and the different error measures that can be used to measure the forecast accuracy. Chapter 4 emphasises the data analysis methods, exploration of data patterns and the findings from these analyses are elaborated. In chapter 5, the research questions are examined based on the analysed findings, while relating back to the previous literature noted in chapter 2. The final chapter 6 concludes with a summary of the main findings and recommendations for aggregation of hierarchical forecasts. It also presents the research contributions and limitations, along with the challenges for future research.

**Chapter 2: Critical review of the literature**

Traditionally, human judgment has played a rather minor role in forecasting developments. However, now there is substantial evidence that forecast accuracy improves when expert judgment is incorporated into the forecasting procedure along with statistical models. Practitioners and academicians have started emphasising more on the blend of judgment with statistical methods rather than on statistical methods alone (Lawrence et al. 2006). Initial forecasts are produced using the statistical software tools that are revised by the domain experts to reflect the effect of special events (Fildes et al. 2009, Syntetos et al. 2009). In this chapter, the previous literature on judgmental forecasting is critically reviewed while covering the various research designs and methods used in this field. The concept of hierarchical forecasts based on different segmentation strategies is illustrated with the help of examples. Different methods used for aggregation of these disaggregated forecasts are discussed further in this chapter. Towards the end of this chapter, the research gaps are identified that are addressed in this study.

**2.1 Judgmental forecasting**

Judgmental forecasting is an active research area with increasing interest over the last 30 years. An extensive literature on determination of the best method that incorporates the judgment factor with the statistical models has been accumulated. It has been found that both human judges and statistical methods are valuable. The combination of expert judgment with the models outperform either decision input in isolation (Blattberg and Hoch 1990). Both the experts (managers) and the models have different strengths and weaknesses. The advantage of using judgment is it brings together a variety of specialised viewpoints (Makridakis and Wheelwright 1989, p. 241). It is most beneficial when adequate historical data is not available for forecasting, for example, there is no historical sales information for new products. It is also useful when the market environment is changing rapidly. At the same time, the decisions based on judgment are intuitive and unaided. Experts are subject to a number of biases like evaluation and perception biases. Sometimes they adjust the SFs without possession of any special event information and mostly do so to have a sense of ownership of the forecasts (Goodwin 2002). As organisations rely too heavily on unstructured judgment, the forecasts are often blurred because of their decisions. The judgmental forecasting process is also very costly (it requires a substantial amount of experts’ time) and sensitive as it disperses responsibility for accurate forecasting (Makridakis and Wheelwright 1989, p. 241). On the other hand, the objective statistical models are unbiased and consistent. The models can take into account many variables and complex relationships (Ord and Fildes 2013, p. 363). However, they produce inputs that are not a complete reflection of reality (Makridakis and Wheelwright 1977). These models require a technical understanding by the decision makers to use it. They come in packages that are expensive and sometimes take a long time to produce the forecasts.

Within the literature of judgmental forecasting, two principle lines of research have emerged. Authors of the first line of research have investigated the forecasting realm using different quantitative methods to determine the best method in terms of accuracy. These methods use the data from different case organisations or experiments to understand the forecasting process when judgment is incorporated into it. In the most recent research line, authors use qualitative methods to capture the dynamics of the judgmental forecasting process. These methods help to discover how the experts use their judgment to make a decision (Stewart 2001). As judgmental forecasting reflects any unnecessary interventions into the forecasting process, organisations are not ready to share this information with academicians. Hence, most of the researches are conducted by quantifying the managerial judgment (Blattberg and Hoch 1990). This is a quantitative approach to qualitative research. Quantitative models are used to understand the process of how individual forecasts are transformed into a single set of group forecasts (Sniezek 1990). These models try to estimate the best model for judgmental aggregation of the experts’ forecasts (Soll and Mannes 2011). Qualitative methods, on the other hand, seek to understand the value and beliefs of how the judgment is applied. Few authors regard the Delphi method to be a subjective and qualitative forecasting method (Firth 1977 p. 204; Gupta and Clarke 1996). Some others think it is an objective way of collecting data from a group of experts (Hannafin 2004). Therefore, the field of judgmental forecasting has been predominated by quantitative research methods. However, qualitative methods can help to understand the human psychology while producing a judgmental forecast. With this understanding, researchers can further develop rules for judgmental forecasting that can help to control the biasness and reduce the forecast error.

Decision makers have tried to understand the judicious mixture of statistical models with human judgment (Hanke and Reitsch 1998, p. 508). The most common approach to forecasting is judgmental adjustments of the initial SFs using expert opinions. Experts have specific domain knowledge about recent events that are not considered by the models and hence they can add extra value to the model forecasts (Fildes et al. 2009). Therefore, this study revolves around this field of judgmental adjustments on SFs. The subsequent sub-sections consider the literature on evaluation of such judgmental adjustments under different categories based on the type of domain information available and research design used. As human judgment is involved while producing the estimates, the amount of domain or contextual information available is very crucial however the amount of accessible information can vary. Much empirical and/or experimental research is dedicated to study the judgmentally adjusted forecasting process. Towards the end of this section, group forecasting techniques have been discussed as a better and alternative forecasting method. Literature based on each of this category is further explored below.

*2.1.1 Unaided and aided judgmental forecasting*

Judgmental forecasting can be of different types based on the available information. The information can be divided into two classes: historical and contextual (domain) information. The historical information reflects the sales history of a product over a period of time. The domain information is all other un-modelled information that can help explaining the past or predicting the future (Lawrence et al. 2006). It comprises of market competition, past and future promotional plans and other environmental changes that could affect the future sales. The statistical models consider historical data and past promotions data whereas the judgmental forecasting process is informed by the non-historical contextual data (Lawrence et al. 2006). There can be two types of judgmental forecasting depending on the type of information available: unaided and aided forecasting. Unaided judgmental forecasting focuses on forecasting a time series with no domain or contextual knowledge. In such a scenario, the judgmental forecaster has access to the same information like the quantitative statistical models (which is most unlikely in practice). As both the methods are restricted to the same set of data, many authors (Lawrence et al. 2006) have studied the comparison between these two methods. The quantitative models outperform the judgmental forecasters (Hogarth and Makridakis 1981) as the judgments are often biased resulting in lower forecast accuracy. On the other hand, in aided judgmental forecasting, the forecasters use the non-time series information to improve the forecasts (Lawrence et al. 2006). The experts are aware of some context associated with the time-series data from their experience. They perform reasonably better than the quantitative models alone (Lawrence et al. 2006) as they know about different special events that had some effect in the past or is likely to have some impact in the future. An increase in the information available in a judgmental setting do not necessarily increase the predictive ability (Makridakis and Wheelwright 1989, p. 250). Nevertheless, domain knowledge is a major contributor of forecast accuracy (Edmundson et al. 1988). When the SFs are revised using relevant contextual information, the forecasts improve drastically (Mathews and Diamantopoulos (1986, 1989, 1990)). Larger adjustments are more advantageous, with smaller adjustments often being ineffective in reducing accuracy (Fildes et al. 2009).

*2.1.2 Empirical and experimental research*

Different authors have explored this field through both empirical and experimental studies. Some of them have used empirical data from different organisations to evaluate the effect of judgment on forecast accuracy. Mathews and Diamantopoulos (1986, 1989, 1990, 1992, 1994) and Diamantopoulos and Mathews (1989) have used the same data from one single company to show that judgmental adjustments did improve the forecast accuracy when they reflect up-to-date information but at the same time, it also introduces bias. Fildes et al. (2009) have used data from four different organisations to analyse the variables responsible for these biases. Negative judgmental adjustments have greater probability of improving the accuracy than the positive adjustments (Syntetos et al. 2010, 2015b). These positive adjustments reflect the optimism bias of the experts (Fildes et al. 2009). However, the benefit of judgmental adjustments depends on the characteristics of the time series and the nature of the adjustments made (Syntetos et al. 2009). Franses et al. (2011) have collected data from the Netherlands Bureau for Economic Policy Analysis (CPB) to analyse pure model forecasts with the expert forecasts. The CPB experts having relevant domain knowledge are able to de-bias the model forecasts. These expert forecasts are better than the model forecasts for shorter forecast horizons. Trapero et al. 2013 have evaluated the effect of judgmental adjustments during promotions using empirical data from a manufacturing company. A hybrid model has been developed to illustrate the fact that experts add value to transfer function models (based on past information). Franses et al. 2009 and Franses and Legerstee (2009, 2010, 2011) have considered data from the same pharmaceutical company to examine the effect of judgmental adjustments on sales data on various parameters. The adjusted expert forecasts explain the actual sales data significantly as the experts possess tacit domain knowledge (Franses et al. 2009). However, the experts deviate too much from the model forecasts and therefore to keep a check on this, proper documentation about the adjustment process should be maintained (Franses and Legerstee 2010). Experts should be aware of the fact that they put too much trust on their own judgment. They make interventions to SFs across all the horizons (Franses and Legerstee 2011) and if they are provided feedback on their past performances, their accuracy improves (Franses and Legerstee 2014). It is also important that the experts understand the inputs and outputs of the statistical models. Syntetos et al. (2010) have proved that improved forecast accuracy from human interventions can reduce the inventory cost significantly. Additionally, inventory implications of adjusting at the forecast level are larger than at the inventory control level (Syntetos et al. 2011, 2015b). Petropoulos et al. 2015 examine the changes in the experts’ behaviour while making judgmental adjustments focusing on cases where there had been big losses due to previously made adjustments. A 29% increase in the probability of another big loss compared to the prior loss probability has been found.

On the other hand, a number of experimental studies have been conducted to analyse the accuracy of the judgmental adjustments. If adjustments are made based on reliable information that are not covered by the statistical models then the forecast accuracy usually improves (Goodwin and Fildes 1999). Generally, forecasters tend to see false patterns in the noise associated with the random fluctuations in the time series and make adjustments to the SFs (Harvey 1995). They have considerable difficultly in underweighting their own judgment relative to SFs even at the cost of accuracy (Lim and O’Connor 1995). There are also studies that combines both experiment and field studies in judgmental forecasting (Goodwin 2000, Legerstee and Franses 2014). Natural experiment with empirical data from case companies have been conducted to see how experts’ behaviour change after providing them various types of feedback (Legerstee and Franses 2014). Based on the empirical results of Franses and Legerstee (2009, 2010) the experts are provided with performance feedback. As a result, the expert deviate less from the SFs, improving the accuracy significantly (Legerstee and Franses 2014). Another benefit of including past performance feedback is of improved interactions between models and experts (Legerstee and Franses 2013). When both SFs and qualitative information (about upcoming promotions) are provided to participants in an experimental setup, they tend to underestimate the promotional effects (Fildes et al. 2015). A consistent bias is seen as previous promotions and SFs act as anchors to the judgmental forecasts being made (Fildes et al. 2015).

*2.1.3 Group judgmental forecasting*

Most of the previous studies have examined individuals making forecasting, while in practice most of the forecasts are made in groups (Makridakis and Wheelwright 1989, p. 250). Individual forecasts have a number of limitations that lead to increased error in forecast accuracy. Group forecasting acts to grab “focus subjects’ attention on those factors promoting “success”, thus encouraging their optimism”(Lawrence et al. 2006). This is also referred to as “*jury of expert opinion”* (Ord and Fildes 2013, p. 367). Group forecasts have been found to be more accurate than combination of individual forecasts (Sniezek 1990). However, group forecasting is subject to a number of factors like social pressure, disagreement and communication error (Sniezek 1990). In hierarchical organisations, strong leaders have considerable influence in the decision-making process. This increases the group pressure for a unanimous opinion (Makridakis and Wheelwright 1989, p. 251). In consensus-seeking group forecasting, members with dissenting opinions feel more pressurized (Makridakis and Wheelwright 1989, p. 252). There are a number of group forecasting methods; three of them are the nominal technique, the consensus method and the Delphi method. The nominal technique is conducted in three steps: first, the group members produce individual forecasts, and then they participate in an unstructured discussion to further revise their forecasts independently (Graefe and Armstrong 2011). The aggregate of the individual forecasts is the group forecast. In the Consensus method, all the members talk to each other face-to-face to decide on a single set of forecasts to which all the members agree. In the Delphi method, the group members have to produce a sequence of forecasts and provide their opinions about the forecasts. The members are kept apart but are informed about the average group forecast (Firth 1977, p. 218). When they arrive at a consensus about the forecasts, it is used as the final forecast. Although this method is highly effective, it can be quite costly and time-consuming (Ord and Fildes 2013, p. 367). When all the group members have access to the same information, the group forecasting method being used have little differential impact (Sniezek 1990).

**2.2 Reconciliation of hierarchical forecasts or cross-sectional aggregation**

Often different divisions within an organisation involve different planning decisions (Ord and Fildes 2013, p. 406). For this reason, the time-series data is often segmented according to various attributes like customer classes, locations or geographies, product-type and SKUs. For example, figure 2.1 shows a three level hierarchy of a clothing company based on product type. The total sales of the company can be decomposed into two sections based on sex: male and female. Each of these can be further divided into finer categories. The sales of male can be disaggregated based on product type into casual and formal wear sales. The sales of female section can be disaggregated into casual, formal and accessories (scarves and hairbands). This results into a hierarchical structure or hierarchical time series (Hyndman and Athanasopoulos 2014).

Figure 2.1: Hierarchy of a clothing company

Ord and Fildes (2013) p.406 describes the concept of parallel hierarchical structure with the example in figure 2.2. Firstly, the total sales is decomposed into different product categories, brands and SKUs. Secondly, the overall worldwide sales can be broken down into countrywide sales and further into within-country regions. Additionally, the daily sales add up to weekly sales, monthly, quarterly and annual sales in a bottom-up fashion. At the most disintegrated level of the hierarchical structure, the sales can be reported at three dimensions viz. SKU, region and time. Each hierarchical structure comprises of a number of individual time-series and their aggregates (Leonard 2014). Due to the ever-changing market environment and competition, managers want to study the segmented data (Weinstein 1987, p. 456). Each individual time-series is examined and different forecasting tasks are performed for each time-series to generate independent forecasts. This results in diverse sets of forecasts across the pyramid structure. Hence, reconciliation of the forecasts, such that sum of the lower-level forecasts is equal to the upper-level forecasts, becomes a necessary and challenging exercise.

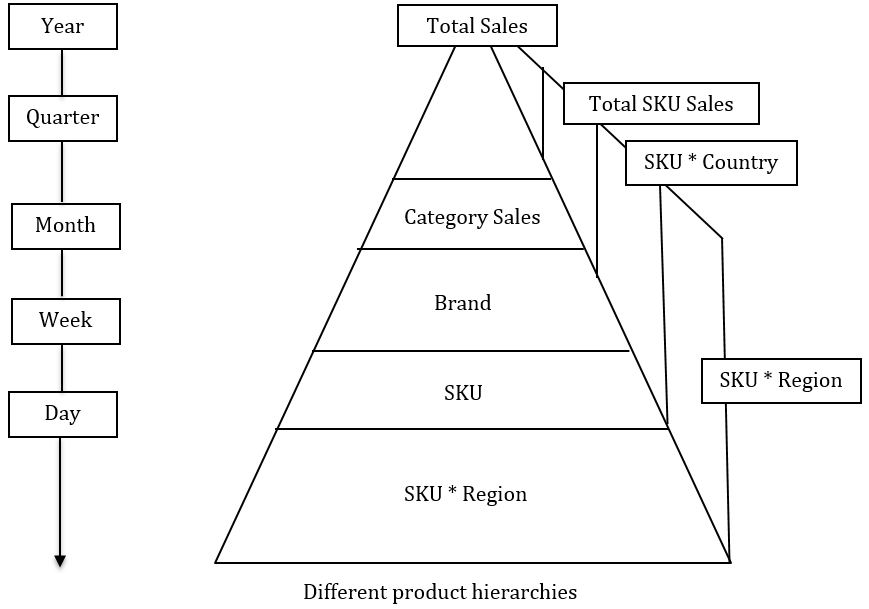


Figure 2.2: Hierarchical attribute of demand data (Ord and Fildes 2013, p. 406)

As the forecasts for each level are generated independently, they do not add up the hierarchy. These independent forecasts have to be reconciled such that the aggregated forecasts are consistent within the structure of the hierarchy (Hyndman and Athanasopoulos 2013b). Cross-sectional aggregation involves aggregation across different attributes like location or product-type. Since Theil (1954), numerous investigations have been reported on the aggregation of hierarchical forecasts. The most commonly used aggregation methods for hierarchical forecasts are the bottom-up and top-down methods (Hyndman and Athanasopoulos 2009). A number of researches have been conducted to compare these two methods (Scharzkopf et al. 1988, Dangerfield and Morris 1992, Fliedner 1999). While some authors believe the bottom-up method to be better (Kinney 1971, Dangerfield and Morris 1992, Zellner and Tobias 2000), others favour the top-down method (Fliedner 1999, Athanasopoulos et al. 2009). Some say there is no advantage of any one method over the other (Ord and Fildes 2013, p.407; Zotteri et al. 2014). There are different situations when one method should be preferred over the other as the relative advantage of one method depends on these situations (Scharzkopf et al. 1988). The middle out method combines both the bottom-up and top-down methods. The optimal combination method proposed by Hyndman et al. 2007 uses a regression model to optimally combine the individual forecasts of all the levels. Each of these four aggregation methods is further illustrated below.

2.2.1 Bottom-up

This is the most commonly used aggregation method in hierarchical time series data (Dangerfield and Morris 1992, Athanasopoulos et al. 2009). Forecasts are produced at the lower levels of the hierarchy and are aggregated upwards to get the upper level revised forecasts (Ord and Fildes 2013, p. 406). Therefore, the upper level forecasts are summations of the appropriate lower level forecasts. The greatest advantage of this method is that there is no loss of information. It helps to capture the dynamics of the individual time-series (Athanasopoulos et al. 2009). However, the disadvantage is that the number of forecasts to be generated depend on the number of nodes in the bottom level. Additionally, it is not easy to model and forecast the bottom level data as it can be quite noisy (Athanasopoulos et al. 2009).

2.2.2 Top-down

This method follows a reverse direction to the bottom-up method. The base forecasts for the top level of the hierarchy are generated and then disaggregated to produce the lower level forecasts of the hierarchy. The advantage of using this method is that only one set of forecasts are to be produced using a single forecasting model. However, there is a huge loss of information in the lower hierarchical levels. The base forecasts are disaggregated using different proportions, which can turn out to be difficult sometimes. As explained by Hyndman and Athanasopoulos 2014, this disaggregation can be done in three ways:

1. *Average historical proportions:*

(1)

where *pj* is the proportion for the *jth* node in the hierarchy. *T* denotes the time horizon which takes values 1,2,….T. *yj,t* is the *tth* observation corresponding to the *jth* node of the hierarchy. *yt* is the observation of the top level at time *t*. Each proportion *pj* captures the average of historical proportions of the bottom level series with respect to the total aggregate *yt* over the period t=1,2,…T.

1. *Proportions of historical averages:*

(2)

Here the notations are same as the previous equation. Each proportion *pj* reflects the average historical proportion of bottom level series with respect to the average total aggregate *yt* over the time t=1,2,….T.

1. *Forecasted proportions:*

This method brings improvements to the historical and static proportions from the previous methods (a) and (b). It produces the independent base forecasts for all the individual series in the hierarchy and then for each level at a time, the proportion of each base forecast relative to the aggregate of all the base forecasts of that level is calculated. These are called the forecasted proportions and using these the top-level forecast is disaggregated to get the revised forecasts for bottom level. This process is repeated for each node going from top to bottom of the hierarchy.

(3)

where *pj* denoted the proportion for the *jth*node of the R-level hierarchy. is the h-step ahead base forecasts of the series that is *l* levels above *j*. is the sum of h-step ahead base forecasts below the series that is *l* levels above node *j* and are directly connected to that series. These forecasted proportions disaggregate the h-step ahead base forecast of the top level data to the h-step ahead revised forecasts of the bottom level data.

2.2.3 Middle-out

This method is a further extension of the bottom-up and top-down methods. Firstly, a middle or intermediate level of the hierarchical structure is chosen. The base forecasts for this level and below are generated. From these base forecasts, the revised forecasts for the all the levels are produced. For the series above this level, the revised forecasts are generated using the bottom-up method on the base forecast. For the series below this level, the base forecasts are disaggregated using the top-down approach. Therefore, this method combines ideas from both the top-down and bottom-up methods.

2.2.4 Optimal combination

Until recently the above three methods were the traditional reconciliation methods used. In this method, the independent base forecasts for each series in the hierarchy are produced. As these base forecasts across the hierarchy do not add up appropriately, reconciliation is necessary. This method optimally combines the independent forecasts to get the revised point forecasts that are reconciled within the hierarchical structure (Hyndman et al. 2011). This method is very flexible, uses all the information available across all the levels of a hierarchy, and allows interactions between series in the same hierarchical level (Hyndman and Athanasopoulos 2014). Another characteristic of this method is if the base forecasts generated are unbiased then the revised forecasts are also unbiased (Hyndman and Athanasopoulos 2013b). The method is explained with the help of following example:

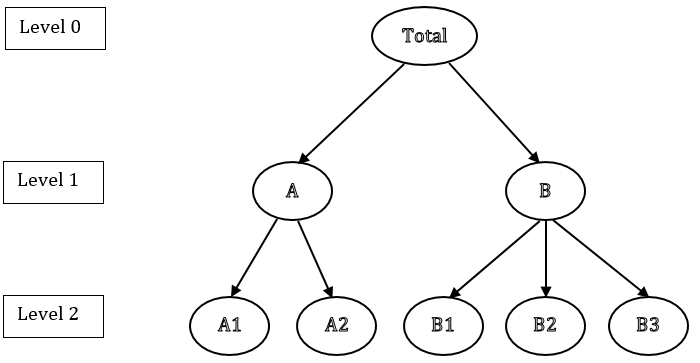


Figure 2.3: A hierarchical structure with three levels

It is a three level hierarchy with two nodes in the intermediate level and five nodes in the bottom level. Let *ni* denote the number of series in the *ith* hierarchical level. The total number of series in the hierarchy will be for an R-level hierarchical structure. Let *S* denote the *n* x *nR* summation matrix reflecting how the bottom-level series are aggregate consistent within the hierarchical structure. *yt*is the vector of all the observations at time *t*. *yR,t* is the vector of all the bottom level observations at time *t*. The above example can be written in matrix notation as:

(4)

This can also be written as

The method is described as it is in Hyndman and Athanasopoulos (2013). In this method, a linear regression model is used to represent the h-step ahead base forecasts.

(5)

where is the vector of h-step-ahead forecasts for the hierarchy

is the mean (unknown) for the bottom level forecasts

is the error of the regression model with zero mean and as the covariance matrix.

With the assumption that (the errors of the regression model approximately satisfy the same aggregation structure as the original data with denoting the bottom level forecast errors); the best linear unbiased estimator for is found to be

(6)

This gives the set of revised forecasts as:

(7)

This method gives more accurate forecasts than the traditional methods: bottom-up, top-down and middle-out (Hyndman and Athanasopoulos 2014). These different aggregation methods for hierarchical forecasting are implemented in the hts package in R (Hyndman and Athanasopoulos 2013b).

**2.3 Research gaps**

As discussed in the section 2.1, judgmentally adjusted SFs results in better forecast accuracy than SFs only. So expert judgment is needed to incorporate all the available information that is not considered by the statistical models. To cope with the high market competition, the managers prefer to study the data at the most disaggregated level and forecast the individual time-series. Sometimes investing in more disaggregated forecasting yields premiums during turbulent periods (Beckenstein 1987). In addition, combination of forecasts is seen as a safer option than individual forecasting methods (Hibon and Evgeniou 2005). Combination procedures can range from statistical methods to using judgment to determine how to combine the forecasts (Lawrence et al. 2006). If managers (experts) make judgmental adjustments to the independent hierarchical forecasts at each echelon of the hierarchy, judgmental reconciliation of the adjusted forecasts through consensus becomes a challenging task. A number of researches have been conducted in the area of cross-sectional aggregation of hierarchical forecasts but all of them explore the statistical reconciliation methods (Athanasopoulos et al. 2009, Hyndman et al. 2011). The judgmental reconciliation of hierarchical forecasts has been identified as a research gap by few authors (Fliedner 2001, Petropoulos 2014, Petropoulos 2015, Spithourakis et al. 2015, Syntetos et al. 2015a).

Management judgment can help to reduce the differences between the forecasts from different aggregation methods (Scharzkopf et al. 1988). Additionally, skilled experts with their subjective judgments can make better forecasts (Kinney 1971). Experts representing different echelons within an organisation have a variety of information needs (Fliedner 1999). With the availability of different useful information, experts make adjustments to the individual SFs at each echelon (node). This has the undesirable consequence of aggregation inconsistency. This means the sum of the lower level forecasts is not equal to the higher level forecast. To make the aggregation consistent throughout the hierarchy, the experts have to reconcile the estimates judgmentally through group consensus. Usually to reconcile the forecasts, adjustments are made in an ad hoc fashion (Hyndman et al. 2011). However, this results in the loss of information and affects the forecast accuracy. On the other hand, to achieve consensus through group interactions especially when judgment is involved is not an easy task. Each of the experts have access to different nature of information about their respective products or divisions. During group interactions, they might have to revise their judgments in the light of new information. This depends upon the vagueness of their previous decision and the reliability of the new information (Goodwin and Wright 2014, pp. 226-228). For example, consider a two-level hierarchy of a movie production company based on geography (figure 2.4). The senior manager responsible for the worldwide income has access to different kind of information compared to the other regional managers. There are many situations when judgmental reconciliation is beneficial as it helps to bring in more information. For example if the sales in Asia is facing competition from another company, this competition might affect the sales of the other regions as well in the future.

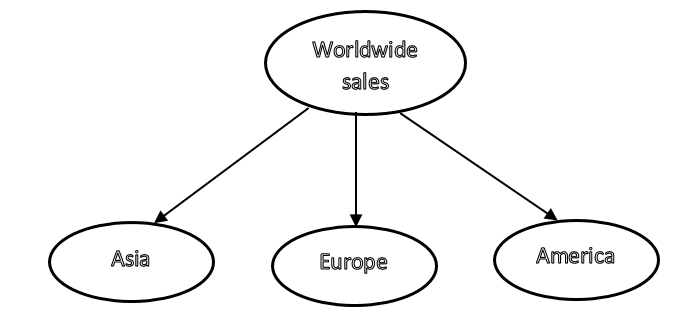


Figure 2.4: A two-level hierarchy based on geographical locations

The judgmental reconciliation process brings together several experts and hence a “larger fund of experience, knowledge and creative insights” (Goodwin and Wright 2014, p. 309). This combination of experts can lead to increase in accuracy as inconsistencies of one judge can cancel out the inconsistencies of the other (Hogarth 1978). The reconciliation process becomes challenging when the opinions and the values of the experts differ. In a business hierarchy, the group decision-making process is more difficult as the presence of powerful individuals inhibits the contribution of those who are lower down in the hierarchy (Goodwin and Wright 2014, p. 319). The purpose of this study is to explore the judgmental reconciliation process in the presence of contextual information with the help of mixed method strategy. The efficiency in terms of forecast accuracy is evaluated. This map directly to the research aim and the research questions outlined in the first chapter.

In this chapter, the literature on judgmental forecasting was critically reviewed. The different research methods used in this field were discussed. The effect of available domain knowledge during forecasting was elaborated followed by empirical and experimental studies in this field. Group forecasting method was suggested as a better and practical forecasting method. Light was thrown on different types of cross-sectional aggregation methods used for hierarchical forecasts. A potential research gap in the direction of judgmentally reconciling hierarchical forecasts was identified. This study attempts to address this gap by conducting a laboratory experiment on forecasting practitioners. The experiment is illustrated in the next chapter along with the different methods used to collect the research data.

**Chapter 3: Research methodology**

This chapter covers the research method employed to answer the research questions outlined in the first chapter. The philosophical position underpinning the research is discussed along with the research design adopted. It also elaborates on the research sample and the data collection methods employed during the study. Finally, the different forecast accuracy measures are illustrated and the justification behind the choice of MAPE and AvrRelMAE as the accuracy measures for this study is explained.

**3.1 Epistemological philosophy**

Epistemology refers to a set of assumptions about the different ways of enquiring into the nature of the world (Easterby-Smith et al. 2012, pp. 18-19). The two main epistemological positions in business management are positivism and interpretivism. The doctrine of positivism being a broad philosophical position is somewhat ambiguous. Positivists believe that knowledge comes from strict scientific methods (Bryman 2012, p. 28). They view the natural world to be independent of the social world. Hence this approach is mainly objective and concepts are defined in such a way that quantitative methodologies can be used to identify and measure them (Guba and Lincoln 2005, pp. 193-194; Easterby-Smith et al. 2012, p. 23). This philosophical position explains human behaviour in terms of cause and effect relationships (May 2001, pp. 10-11). From observations on social phenomena, the researcher can make generalisations about the population as a whole (May 2001, p. 10). A deductive approach is applied involving the development and testing of hypotheses (Guba and Lincoln 2005, pp.193-194). Interpretivism, on the other hand, is a contrasting position to positivism. It respects the differences between the natural and social world (Bryman 2012, p.30). Social reality is considered to be a product of different interpretations that people can associate with it and the social scientist needs to grasp the subjective meaning of the social action (Saunders et al. 2007, p. 116; Bryman 2012, p. 30; Blumber et al. 2012, p. 19). This position employs qualitative methods to gather data from which ideas are induced (Easterby-Smith et al. 2012, p. 24). Positivists believe that the knower and the known are independent while the interpretivists regard them to be inseparable (Teddlie and Tashakkori 2009, p. 85).

The research field of forecasting has been predominated by the positivist paradigm. Mostly forecasters look for combination of various variables that can help improve the forecast accuracy. They explore causal relationships between different variables that have an effect on the forecast accuracy. The research is undertaken in a value-free way i.e. the observer is independent from the research settings (Easterby-Smith et al. 2012, p. 24). Secondary data is collected from different organisations and quantitative methods are used to derive insights from them (Lawrence et al. 2006). Hence, the researchers embrace an objective position. However, in the field of judgmental forecasting, qualitative methods are required to discover how experts use their judgment to make a forecast or adjust an SF. As explained earlier, it is difficult to collect qualitative data when it is related to human judgment. The Delphi method is often argued to be a qualitative forecasting method of achieving consensus among a group of experts (Gupta and Clarke 1996). It is a structured and effective method of group communication process, which helps in harnessing experts’ opinions over a series of survey rounds (Gupta and Clarke 1996). As the survey questions are in form of numbers and figures, some support the fact of it being a quantitative research method of processing data (Landeta and Barrutia 2011, Easterby-Smith et al. 2012, pp. 24-25). There is a long-standing debate about the epistemological position underpinning the Delphi method. As the researcher’s position is uninvolved and objective, the method aligns with the assumptions of the positivistic paradigm (Robson 2002). It follows the deductive way of constructing data embracing the ontological position of one single reality, which the experts agree upon (Hanafin 2004). On the other hand, it also follows the interpretivist philosophy with social constructionism as the ontological position. As the contribution of each participant is recognised and acknowledged, it fits in to the definition of social constructionism (Lincoln and Guba 1985, p. 82). The opinions and views of the experts are taken into consideration, which makes it subjective in nature. Therefore, it is difficult to come to a strong conclusion about the epistemological positions behind forecasting methods in general.

This research is a blend of both positivist and interpretivist epistemologies. It examines the correlation between judgmental adjustments and forecast accuracy through causal relationships. As discussed in the second chapter, judgmental adjustments to SFs help to make better forecasts and this drives the process of data collection for this study. This deductive nature of data collection method points towards a positivist paradigm. The interpretivist position is supported by the fact that qualitative methods like focus groups are used to collect the data. Subsequently, as insights from the collected data influence the theory behind this study, it also follows an inductive reasoning process. Therefore, it is a mixed method research where the two research strategies, quantitative and qualitative, are combined to understand the logic to theorize observations. It uses a combination of both inductive and deductive logic in a distinctive sequence called the “inductive-deductive research cycle” (Teddlie and Tashakkori 2009, p. 26). This research cycle as depicted in figure 3.1, can be seen as moving from grounded facts to general inferences through inductive inference and from general inferences to predictions through deduction inference (Teddlie and Tashakkori 2009, pp. 26-27). Depending on the purpose of the study, the induction could come first or the deduction could come first in this research cycle. At any given time, research relating to any given question falls somewhere within this inductive-deductive research cycle (Teddlie and Tashakkori 2009, p. 87).

Generalizations, Abstraction, Theory

Prediction, Expectation, Hypothesis

Deductive reasoning

Inductive reasoning

Figure 3.1: Inductive-Deductive research cycle (Teddlie and Tashakkori 2009, p. 27)

As the key aim of this study is to understand the rationale behind the judgmentally adjusted hierarchical forecasts using mixed method strategy, it will adopt a combination of positivism and interpretivism positions. The pragmatism paradigm embraces features associated with both these positions. Business forecasting is often teamed with the principles of pragmatism (Ord and Fildes 2013, p. xvii). Pragmatism holds that the current state of the research question drives everything for mixed-method approaches (Saunders et al. 2007, p.122; Teddlie and Tashakkori 2009, pp. 22-23). It combines both inductive and deductive logic (Teddlie and Tashakkori 2009, p. 22), and it is considered advantageous to do so (Saunders et al. 2007, p.119). The social entities are believed to exist both objectively and subjectively through inter-subjectivity. This implies that there is one single reality but the actors interpret it in different ways. Cherryholmes (1992, pp. 13-14) cited in Teddlie and Tashakkori 2009, p. 90 describes pragmatists as follows:

*For pragmatists, values and visions of human action and interaction precede a search for descriptions, theories, explanations, and narratives. Pragmatic research is driven by anticipated consequences… Beginning with what he or she thinks is known and looking to the consequences he or she desires, our pragmatist would pick and choose how and what to research and what to do.*

A paradigm contrast table for the three philosophical positions: interpretivism, positivism and pragmatism is presented in table 3.1. It has been argued by many authors that this integration of two different epistemological philosophies is incomparable and incompatible. However, others do agree that there are areas of overlap and communality between the paradigms (Bryman and Bell 2007, pp. 643-644). As May (2001) p. 7 observes, “these perspectives do not determine the nature of the research process itself for there is constant interaction between ideas about the social world and the data collected on it.”

|  |  |  |  |
| --- | --- | --- | --- |
| ***Dimensions of Contrast*** | ***Interpretivism*** | ***Positivism*** | ***Pragmatism*** |
| Methods | Qualitative | Quantitative | Both qualitative and quantitative; researchers answer questions using best methods |
| Logic | Inductive | Deductive | Both inductive and deductive |
| Epistemology (researcher/participant relationship) | Subjective point of view; reality co- constructed with participants | Objective point of view | Both objective and subjective points of view, depending on the stage of the research cycle |
| Axiology (role of values) | Value-bound inquiry | Value-free inquiry | Values important in interpreting results |
| Ontology (The nature of reality) | Ontological relativism- multiple, constructed realities | Naïve realism (an objective, external reality that can be comprehended) | Diverse viewpoints regarding the social entities; best explanations within the personal value systems |
| Possibility of causal linkages | Impossible to distinguish causes from effects; credibility of descriptions important | Real causes temporally precedent to or simultaneous with effects | Causal relations, but they are transitory and hard to identify; both internal validity and credibility important |

Table 3.1: Paradigm contrast table (adapted from Teddlie and Tashakkori 2009, p. 88)

**3.2 Research design**

The research design is a framework for the data collection, measurement and analysis of the research process (Cooper and Schindler 2006, pp. 138-137). As described by Ghauri and Grønhaug (2002) p. 47, “the research design is the overall plan for relating the conceptual research problem to relevant and practicable empirical research”. It acts like a blueprint that helps to select an approach that answers the research problem in the best possible way considering the given constraints (Ghauri and Grønhaug 2002, p. 47). The different types of research designs are *experimental design*, *cross-sectional design*, *longitudinal design*, *case-study design* and *comparative design*. An experimental design is used to test the causal relationships between different independent and dependent variables. It consists of random assignment of subjects to either the experimental group (group exposed to the experimental stimulus) or the control group (group not exposed to the experimental stimulus) (Easterby-Smith et al. 2012, p. 40). The characteristic of this design is the researcher is allowed to manipulate the different experimental conditions in order to assess their effects (Ghauri and Grønhaug 2002, pp. 52-53). Researchers can learn easily from their experiments that helps in designing and conducting new and better experiments. This enables a better learning approach to empirical research (Siemsen 2011). Researchers in the field of forecasting prefer using laboratory experiments as these help to study the subjects’ behaviour in probabilistic and deterministic environments. In addition, they prefer conducting experiments to empirical data as it is difficult to get access to judgmental forecasting methods in organisations. In the field of logistics and supply chain, the experimental design has the potential of informing a number of research questions (Deck and Smith 2013). Because of the reliance on the observation of behaviour and on the analysis of data, behavioural laboratory experiments are also classified as empirical methods (Siemsen 2011). For the purpose of this study, a laboratory experiment was designed to see how consensus is achieved through group interaction while estimating the future.

*3.2.1 Experimental design*

A hypothetical situation was created considering a two-level business hierarchy of the company *Apple*. *Apple* was chosen as the case company as sales data are made available to the public after each quarter. In addition, *Apple* being one of the most popular companies in the UK, the participants could use their own experiences and knowledge (other than provided information) to make their decisions. Three products (*iPhone*, *iPad* and *Mac*) of the company were chosen as the lower hierarchical level. These three products were the top three players of *Apple* with respect to market share of the company. Also there was inconsistency in the sales figures of the other products (sometimes combined with accessories) in the quarterly sales reports. The upper level represented a senior position of *Apple* who had the three products under her/his authority. The participants were distributed into groups of four and randomly allocated to one of nodes of the 1-3 pyramid as shown in figure 3.2. So three members of each group represented three product managers corresponding to the three products: *iPhone*, *iPad* and *Mac* respectively, and the fourth member was the senior manager of *Apple*. As the aim of the experiment was to see how group consensus is achieved when judgmental interventions are linked to hierarchical forecasts, focus group was chosen as the qualitative method to gather in-depth information relating to the feelings and opinions of the participants (Collis and Hussey 2009, p. 155). As suggested by Krueger and Casey (2015), p. 82 that more insightful data is best achieved in smaller focus groups, the number of members in each group was decided to be four representing the four nodes of the 1-3 pyramid.

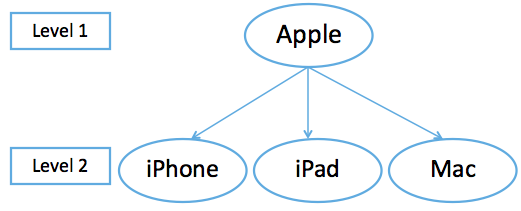


Figure 3.2: A two-level business hierarchy of *Apple*

The participants were provided with the following data:

1. **Statistical forecasts (SFs):** The historical sales (quarterly) figures for the three products were collected from the *Apple* company website (Apple 2015). The data was collected from quarter 4 of 2011 to quarter 4 of 2014. This time frame was decided as there were inconsistencies in the information released in the website before quarter 4 of 2011. The sales figures for the *Apple* senior manager were the addition of sales of three lower level products for each quarter. The common statistical methods used to produce forecasts are the naïve method, moving averages, Simple Exponential Smoothing (SES), Holt’ exponential smoothing and Holt-Winter exponential smoothing. These five methods have been explained in the table 3.2.

|  |  |  |
| --- | --- | --- |
| *Method* | *Mathematical representation* | *Characteristics* |
| Naïve method | Ft+i = forecast for period t+i  t = present period  i = number of periods ahead being forecast  Xt = actual value for period t | The most recent available information is used as a forecast.  It cannot accommodate randomness. |
| Moving averages | Ft+1 = forecast for period t+1  i = present period  N = number of values included in averages  Xi = actual value for period i | Average of a number of observed values is taken as forecast for the next period.  It allocates same weightage to all the values. |
| Simple exponential smoothing (SES) | ,    Ft+1 = forecast for period t+1  t = present period  N = number of observations  Xt = actual value for period t  Ft = forecast for period t  = data smoothing factor | It gives most weight to the recent observation and decreasing weights to the older values.  It is used for data without any trend or seasonality. |
| Holt’s exponential smoothing | Ft+m = forecast for period t+m  St = equivalent of SES smoothed value  = trend smoothing factor  Tt = smoothed trend in data series  Tt-1 = last smoothed trend  m = number of periods ahead being forecast | It considers the trend factor of the data.  It does not consider the seasonality in data. |
| Holt-Winter exponential smoothing | S= smoothed value of deseasonalized series  = seasonality smoothing factor  T= smoothed value of trend  I = smoothed value of seasonal factor  L = length of seasonality (like months or quarters) | It deals with seasonal data series with trend. |

Table 3.2: Description of statistical forecasting methods

Examining the historical time-series data for each product, the best statistical method among these five statistical methods was used to obtain the SFs for the four quarters of 2015 in R (R code is presented in Appendix A). The statistical method used for the generation of the SFs deferred for each product as the time-series data for each product had different patterns. From figure 3.3, it can be observed that the historical data of *iPhone* and *Apple* show linear (upward) trends therefore, the Holt’s exponential smoothing (with damped trend) method was applied to produce the SFs for these two products. Simple exponential smoothing method was used for *iPad* and *Mac* as their historical data do not show any trend. Seasonality was not taken into account for any of the products to avoid consideration of the past product launch periods as a seasonal pattern. For example, the new *iPhone* versions have been launched in the third quarter for the last couple of years but it may not be so for 2015. The SFs for 2015 along with the historical sales data were presented to the participants in the form of graphs (Appendix G). Harvey and Bolger (1996) had found that judgmental forecasters perform better when the data was presented in graphs than in tables (provided the data was trended). From the figure 3.3, it can be seen that the *iPhone* and *Apple* historical sales data have an upward trend. To maintain consistency among the group members, graphs were chosen as the common display format for all the products.

Figure 3.3: Historical sales of *Apple*, *iPhone*, *iPad* and *Mac*

1. **Contextual or domain information:** Each manager was provided with some qualitative information, in form of news articles or blogs from different web sources, regarding anticipated sales of their respective product over 2015. To maintain homogeneity, each manager was provided with 3 pieces of information. The information shared was mutually exclusive among the managers. Therefore, each manager had different information with respect to the others.

*3.2.2 Research sample*

The subjects chosen for this experiment should possess forecasting knowledge: forecasting practitioners or forecasters. Given the time constraints and duration of the study, it was difficult to get access and consent from real managers to participate in such an experiment. Masters students and young researchers (PhD and staff) from LOM section were invited to participate in the experiment. Homogeneity among the focus group members was maintained to encourage more in-depth and open discussions. Seven focus groups (labelled A-G) were conducted with four participants in each group. Two groups (A, B) were part of a pilot study and the findings from the five main groups (C-G) are reported in this thesis.

There has been a lot of speculation about the validity of using student samples in many social-science disciplines like management (Bello et al. 2009). Many scholars argue that results obtained from student samples are not generalizable to the wider population of professionals (Highhouse and Gillespie 2009). On the other hand, some agree that logistics students are appropriate participants for decision-making experiments. Thomas (2011) explains, “the students that have received training in a speciﬁc area of logistics are suitable candidates for potential participation in experimental research in that area of logistics”. For an experimental research design, maintaining the homogeneity among the participants is very important (Thomas et al. 2010). Students are considered as the most homogenous population available for experimental research (Thomas 2011). Another advantage of using students is they do not have the baggage associated with being with an organisation (Deck and Smith 2013). Many authors have conducted forecasting experiments using student samples (Lawrence et al. 1985, Sniezek 1990, Graefe and Armstrong 2011, Spithourakis et al. 2015). As explained in section 2.1, judgmental forecasts being highly sensitive information, organisations are not ready to share such information with academic researchers. So had the experiment been conducted with real managers, the managers would respond to a hypothetical situation as well. It has been observed that when professionals are taken out of their familiar organisational context, they perform no better than the students (Deck and Smith 2013). Daniel Kahneman and Amos Tversky conducted a number of small, artificial experiments on student samples to see how people make uncertain financial decisions. Their work on behavioural economics helped Kahneman garner the 2002 Nobel Prize in Economics \* (Highhouse and Gillespie 2009). This is an excellent example of how work on student samples can have real world relevance.

*3.2.3 Data collection methods*

The task involved predicting the quarterly sales of the three products (*iPhone*, *iPad* and *Mac*) and the overall *Apple* level for the year 2015. The task was divided into two rounds:

1. **Individual judgment round (questionnaire):** The task material contained a questionnaire corresponding to the product each manager represented (Appendix G). The questionnaires included time-series graphs of historical sales of the respective product from quarter 4 of 2011 to quarter 4 of 2014. The graphs also had the SFs for the four quarters of 2015. The questionnaires had contextual information about each product and the source of each piece of information was attached as appendices. The participants had judgmentally adjust the SFs for each quarter of 2015 using the contextual information. They were asked to provide the adjusted forecasts and specify their rationale behind the judgments (optional). It was the participants’ choice if they wanted to adjust the SFs. They could make no adjustment and go forward the SFs if they considered them to be reasonable. The characteristic of this round was that it was an individual round and so the participants had to apply their own judgment regarding the adjustment without discussing with other members of the group.
2. **Group judgment round (focus group or consensus group technique (Sniezek 1990)):** In this round, all members of a group had to reconcile the independent forecasts of the previous round such that the forecasts of the lower level add up the hierarchy. This was to be achieved through group consensus, which meant that for each quarter the aggregate of the forecasts for the products *iPhone*, *iPad* and *Mac* should be the same as the forecast of the upper hierarchy *Apple*. As judgmental

\* Kahneman received the Nobel Prize six years after Tversky’s death and as this is not awarded posthumously, Tversky did not receive the award.

adjustments were made during the individual round, the forecasts usually did not add up. Therefore, the group members had to discuss among themselves the rationale behind each adjustment made and how reasonable that adjustment was. Each participant (representing different products) had different contextual information about the expected market changes. There might be facts where the sales of one product was being affected by the changes brought about in the other products’ market. For example, there was a rumour about 13’’ *iPad* making an entry to the market. This might have some effect on the *Mac* sale figures that the *Mac* managers were oblivious about. The group discussions allowed them to examine some of these facts. The managers had to readjust their individual forecasts to reconcile them in the light of new information. They could reconcile in any way: bottom-up, top-down or by making adjustments at both the levels. This round was audio-recorded and participant observation notes were taken for analysis.

At the end of the second round, the participants were asked two oral questions:

1. Who do you think has more influence in the decision making process? The top level of the hierarchy or the bottom level?
2. Which forecasts are you more confident about: individual or group forecasts?

As mixed method data collection strategy is employed, it helps to integrate the quantitative and psychological aspects of decision-making (Goodwin and Wright 2014, pp. 3-4). This study follows the parallel mixed design (figure 3.4) where the qualitative and quantitative strands occur simultaneously or with some time lapse (Teddlie and Tashakkori 2009, p. 26). The qualitative and quantitative phases are planned and implemented to answer the same research question (Teddlie and Tashakkori 2009, p. 22). This is also referred to as methodological triangulation where multiple methods are used to study the same problem. Triangulation is the combination of more than one method or data source to study a social phenomenon (Bryman 2012, p.717). The results from an investigation employing one strategy can be cross-checked against the results of another strategy (Bryman 2012, p. 635). As the field of judgmental reconciliation of hierarchical forecasts is very novel, mixed methods aid in understanding a person’s behaviour or experience in this context. This helps to delve into the meanings behind the adjustments made to SFs. Focus groups help to gather rich in-depth data from the respondents, in their own expressed words and reactions. A distinctive characteristic of this method is the involvement of the researcher is minimum. This acts both as a strength and as weakness. Strength as the influence of the researcher is very less. At the same time, the researcher should carefully design the structure so that the discussion does not drift away from the chief issue. It is also a flexible, convenient and inexpensive way of gathering information from several respondents in a short time (Ghauri and Grønhaug 2002, p. 109). This method generates a huge amount of qualitative data that is sometimes difficult to manage and categorise. Additionally, the participants can understand the experimental task in different ways that might affect their judgments (Goodwin and Wright 2015, p. 263). Nevertheless, mixed method “enables researchers to simultaneously ask confirmatory and exploratory questions, thus verifying and generating theory in the same study” (Teddlie and Tashakkori 2009, p.152). A comparison between the three different positions: qualitative, quantitative and mixed methods is presented in table 3.3.

Conceptualization stage

Experiential Stage (Methodological)

Experiential Stage (Analytical)

Inferential Stage

Meta-Inference

Figure 3.4: Graphic illustration of parallel mixed designs (Teddlie and Tashakkori 2009, p.152)

|  |  |  |  |
| --- | --- | --- | --- |
| ***Dimension of Contrast*** | ***Qualitative Position*** | ***Quantitative Position*** | ***Mixed methods Position*** |
| Methods | Qualitative methods | Quantitative methods | Mixed methods |
| Paradigms | Constructivism  (and variants) | Postpositivism;  Positivism | Pragmatism; transformative perspective |
| Research questions | QUAL research questions | QUAN research questions | MM research questions (QUAN plus QUAL) |
| Form of data | Typically narrative | Typically numeric | Narrative plus normal |
| Logic | Inductive logic | Hypothetico-deductive | Both inductive and deductive |
| Research designs | Ethnographic research designs and others (case study) | Correlational; survey; experimental; quasi-experimental | MM designs, such as parallel and sequential |
| Sampling | Mostly purposive | Mostly probability | Probability, purposive, and mixed |
| Data analysis | Thematic strategies; categorical and contextualizing | Statistical analyses: descriptive and inferential | Integration of thematic and statistical |

Table 3.3: Dimensions of contrast between three methodological positions (Teddlie and Tashakkori 2009, p. 22)

*3.2.4 Ethics*

As the experiment involves human subjects and their judgments, ethics approval was taken before inviting the subjects to participate. After gaining approval from the ethics committee, invitation letters were distributed among the subjects requesting them to participate in the study. The invitation letter did mention that the anonymity of the participants would be maintained. Written consent was taken from the all the participants before starting the experiment that included their consent for audio recording the group round. Each participant was given a £5 amazon gift voucher as a token of appreciation for their participation. A scanned copy of the approved ethics form is attached in Appendix F.

*3.2.5 Pilot study*

A pilot study was conducted before the main research with two focus groups as a part of the qualitative taught module (BST214). A few changes in the main study were made after the pilot. Firstly, a time limit was allocated to the groups for completion of both the individual and group rounds. During the pilot study, there was no time restriction and the groups spent more than 2.5 hours to complete both the rounds. Even in the organisations, the decision making process has to be completed within a pre-determined time frame. Therefore, a time limit of 180 minutes was decided for the two rounds. Secondly, since both the groups mostly opted for the bottom-up approach, the average historical proportions corresponding to each product was provided in the group round questionnaire to encourage the participants to consider top-down as an option for reconciliation. Finally, the time-series graphs in the questionnaires were modified by explicitly mentioning the numbers for each data point.

**3.3 Accuracy evaluation**

As the old saying goes ‘*the proof of the pudding is in its taste*’, one needs to check how well the forecasting method performed (Gross and Peterson 1983, p. 68). No one method can provide the best forecast, so the forecasts need to be evaluated to check for their accuracy. To evaluate the forecasts, their accuracies are calculated. Many authors refer to accuracy as “goodness of fit” (Makridakis and Wheelwright 1989). Forecast accuracy is very important for planners and managers. It enables them to make better decisions and help to improve the efficiency and effectiveness of the supply chain (Zotteri et al. 2014). There are different types of accuracy measures for different applications but the goal is common- to minimise the forecast error (Mahmoud 1987).

Thus, the error term is defined as

**Error = Actual – Forecast**

(8)

or **et = Yt - Ft**

where *et* is the error at time period *t*

*Yt* is the observed or actual value at time period *t*

*Ft*is the forecasted value for time period *t* produced at the end of some past time period (t-1, t-2…).

If *et*is positive, it is under-forecasting where the actual value is greater than the forecast. If *et*is negative, it is over-forecasting where the actual value is lesser than the forecast. The smaller the error term is, the closer the forecast is to the actual value.

A summary of different types of error measures (with *n* number of observations):

1. Mean Error (ME)

This is the average of individual errors corresponding to the n observations. This is simplest error measure to calculate. It helps to detect bias in the forecasts. The ME will be positive (negative) when the actual value is greater (lesser) than the forecast value (Ord and Fildes 2013, p. 44). However, the positive errors can cancel out the negative errors resulting in a zero mean error.

(9)

1. Mean Absolute Error (MAE) or Mean Absolute Deviation (MAD)

This is the average of the magnitudes or absolutes of the forecast errors. It does not take into account the positive or negative sign of the error. This is helpful in avoiding having an average error close to zero. This is useful when one needs to measure the accuracy in the same units as the original data (Hanke and Reitsch 1998, pp. 110-113). This is usually greater than the ME and can attract the management’s attention. It puts equal weight on the level of error (Firth 1977, p. 34) which is sometimes not necessary.

(10)

1. Mean Absolute Percentage Error (MAPE)

This calculates the errors in terms of percentages rather than amounts. This approach is useful when the size or magnitude of the error compared to the original data is important (Hanke and Reitsch 1998, pp. 110-113). MAPE has the advantage of being scale independent and therefore is useful for comparing the accuracy across different data sets or between different forecasting methods (Hanke and Reitsch 1998, p. 113). It is the best choice for positive data that are much greater than zero (Hyndman and Koehler 2006).

(11)

There are two disadvantages of using MAPE.

* If the actual value *Yt*is zero, than the divisor becomes zero. Hence, *Yt* should be greater than zero.
* The percentage errors assume a meaningful zero that on some scales like temperature does not make any sense.

1. Mean Squared Error (MSE)

The error terms are squared, summed and added to obtain MSE. However, this approach gives a very high penalty for forecasting errors as these are squared. The forecasts with large deviations are penalised more than those with smaller ones. It is one of the most commonly used error measures and is analogous to the statistic variance ().

(12)

1. Root Mean Squared Error (RMSE)

This brings MSE calculation to the same units as the original data. For example, if the data is in Dollars ($) than MSE will result in dollar squared. RMSE restores the original units of data and hence making interpretation more straightforward (Ord and Fildes 2013, p. 46). However, it gives more weight to large (absolute) errors. It is analogous to the statistic standard deviation ().

(13)

1. **Average Relative MAE (AvgRelMAE)**

Davydenko and Fildes (2013) have introduced this new accuracy measure. It uses the geometric mean of MAE ratios to define an average relative MAE measure. Considering the SFs as the baseline forecasts, AvgRelMAE shows how much the forecast accuracy is improved or reduced by the judgmentally adjusted forecasts. It has the advantage of easy interpretation. If AvgRelMAE < 1 than on average and therefore, judgmental adjustments improve the accuracy. Whereas AvgRelMAE > 1 indicates the opposite. This measure is informative and uses all available information effectively (Davydenko and Fildes 2013). It represents the performance of the judgmental adjustments objectively without introducing any subjective bias. The only limitation of this measure is that the errors or MAEs have to be non-zero.

(14)

Here is the MAE of the baseline SFs for time series *i* and is the MAE of the judgmentally adjusted forecasts for series *i*. *m* is the total number of time series and *ni* is the number of non-zero errors for judgmentally adjusted forecasts for series *i*.

(15)

are the errors of SFs and judgmentally adjusted forecasts respectively for period t and series *i*, and *Ti* is the set of time periods for which are available (non-zero).

All forecast accuracy measures have some advantages and disadvantages. The four *Apple* data sets have different range of data, so scale free measures have to be considered to make comparisons across the series. As *Apple* sales data are all positive numbers, MAPE can be used to compare the individual forecasts with the group reconciled forecasts. A number of authors have used MAPE as the accuracy measures in their researches (Fildes et al. 2009). Moreover, there are two methods of forecasting in the experiment: individual judgmental and group judgmentally reconciled forecasts. To compare the judgmentally adjusted forecasts with the control group forecasts which are the statistically aggregated forecasts from the individual forecasts, AvgRelMAE can be used.

This chapter covered the methodological foundations underlying this study. The philosophical underpinnings of this study were explained along with the research design. The laboratory experiment to see how hierarchical forecasts are reconciled judgmentally through group consensus was elaborated. The advantages and disadvantages of using the data collection methods were discussed. The different accuracy measures in forecasting were covered along with the strengths and limitations of each measure. The next chapter describes the data analysis methods employed for analysing the quantitative and qualitative data collected.

**Chapter 4: Analysed findings**

In this chapter, the results from the laboratory experiment are presented. The experimental group for this research consists of the participants’ group forecasts. The control group should (typically) comprise of benchmark forecasts against which the experimental group forecasts can be compared. However, the purely SFs cannot be the benchmark as they do not account for the impact of the forthcoming special events, such as launches of new product versions. Furthermore, the individual non-reconciled forecasts from the participants cannot be considered as the baseline forecasts. Although these individual forecasts may be more accurate than the group forecasts, they are not reconciled. Statistically reconciled forecasts from different statistical aggregation methods described in section 2.2 can be regarded as the benchmark. Hence, the aggregated SFs generated using the individual participants’ forecasts (before reconciliation) is the control group of this experiment, with whom the group reconciled forecasts are compared.

Being a mixed method research, the data analysis process uses an integration of thematic and statistical methods (Teddlie and Tashakkori 2009, p. 22). This parallel mixed data analysis involves two separate processes: the qualitative data analysis using thematic analysis of the narrative data, and descriptive or inferential statistical analysis of the quantitative data (Teddlie and Tashakkori 2009, p. 266). The qualitative data from the focus groups and observations made during the experiment are explored to identify themes and categorise the results (Krueger and Casey 2015, pp. 118-119). Forecast accuracies are calculated for the forecasts obtained from the participants across different dimensions like time horizon (quarter 1, quarter 2 and both), product type and hierarchical level. Deviations from the individual forecasts to get the group forecasts are computed for detection of any direct relationship between these two sets of forecasts. Inference based on the results from both the qualitative and quantitative strands are integrated to form meta-inferences about the reconciliation process (Teddlie and Tashakkori 2009, p. 152). Meta-inferences are conclusions based on findings obtained from both the qualitative and quantitative components. The discussions based on these conclusions are outlined in the next chapter.

**4.1 Qualitative data analysis**

Qualitative data analysis is the analysis of various types of narrative data in the form of audio, videos or text (Teddlie and Tashakkori 2009, p. 251). It typically follows an inductive approach to research (Teddlie and Tashakkori 2009, p. 22) and leads to generation of themes that are grounded in the data (Lincoln and Guba 1985, p. 344). In the laboratory experiment, the focus group conversations (audio-recorded with informed consent from the participants) and the observation notes provided rich qualitative data (Krueger and Casey 2015, pp. 6-7). Additionally, the participants presented their rationale behind each judgmental adjustment of the individual round in textual form in the questionnaires. These textual data along with the recorded data are analysed to identify and trace the dominant themes within the data. The classic analysis strategy is used for distinguishing the themes (Krueger and Casey 2015, pp. 118-121). The transcripts and observation data are colour coded to represent the five groups. The participant quotes are cut out and placed according to different categories. The same process is followed for the observation notes. These different categories are rearranged or combined depending on their frequencies, emotions and extensiveness. A contrast comparison is made among the groups based on the synopsis of these categories. This leads to the determination of the dominant themes associated with judgmentally reconciled SFs (figure 4.1):

Figure 4.1: Representation of the four themes

1. Authority:

The decision making process in an organisation is reflected within the different hierarchical levels. The most prominent theme in hierarchical decision making is the leadership influence or authority. According to Anderson and Brown (2010), “hierarchies pervade social groups and have a profound impact on group functioning.” In group decision making, consensus is a result of interaction in the form of open communication (Harrison 1975, p. 191). The presence of an influential individual can inhibit the performance of those in lower hierarchical levels (Goodwin and Wright 2014, p. 319). This can prevent the expression of critical ideas, which was observed from the focus group discussions. Even when the product managers (lower level) were more confident about their forecasts, the senior manager made the final decision for the reconciled forecasts after heeding their arguments. There were times when the product managers were not quite happy with the senior manager’s decision but had to follow it because of the hierarchical difference. In one group, the *iPad* and *Mac* managers did not agree with the judgmental forecasts of the *iPhone* manager. They asked the *iPhone* manager to reconsider her/his forecasts for the reconciliation process highlighting some reasonable assumptions. However, the senior manager had confidence on the *iPhone* manager’s judgments and asked the other two managers to readjust their forecasts for reconciliation. This shows that the upper hierarchy has a major influence in the decision making process. The effect of hierarchy also attributes to the shifting of suggestions during the reconciliation process. Moreover, sometimes the upper hierarchical level wants to have ownership of the forecasts and they try to dominate the others to achieve it. Therefore, the experts or managers are highly influenced by the organisational politics and are under social pressures to achieve consensus (Blattberg and Hoch 1990). Thus for effective group decision making, the leader should not be dominant and there should not be a struggle for power (Gilligan et al. 1983, p. 53).

1. Expertise:

A considerable amount of expertise is required in organisations to develop a procedure for handling the strategic decision making process. This is reflected from the focus group discussions where the product managers (*iPhone*, *iPad* and *Mac*) did influence the reconciliation process. Moreover, the rationale provided by the participants behind their individual adjustments also direct towards their expert skills. Within this theme, there are two subthemes: market knowledge and experience. The product managers have more information about their respective products than the senior management. They are the experts in their divisions and hence did have more detailed knowledge about their respective products. They know how different updates in design and function of one product can affect its sales. They have a substantial amount of experience in sales forecasting that guide their managerial intuitions regarding the forecast decisions. For example, the *iPhone* managers did consider the fact that if the production of *iPhone* *5C* version is discontinued (this was given as a piece of contextual information in the questionnaire) than there will be few people who would like to buy it because of the reduced price. In another group, the *iPhone* manager considered the rising interest in selfies as a good reason for increase in sales of the new camera improved *iPhone* version. From their experience, these managers (including the senior manager) know about the seasonality factor in time-series forecasting. The statistical methods used to produce the SFs did not take into consideration the seasonality aspect of the historical data. The sales of all the products were highest in the fourth quarter of every year. The main reason behind the high sale figures is the festive season of Christmas and New Year, and a number of discount offers (Black Friday and Boxing Day). Although this factor was not covered by the statistical methods, it was taken into account by the managers while making adjustments to the SFs. As majority of customers wait for reviews about the new products before buying them, it affects the new product sales. Some of the product managers were able to interpret this fact as well. Few managers even included the instability of demand-supply equilibrium for new products while making judgments. Therefore, the experts diagnose the past data and future expectations, and then estimate the forecasts (Blattberg and Hoch 1990).

1. Product:

The contribution of managers in the reconciliation process was determined by their respective products. The decision making task is driven by different characteristics of the product such as its current stage in the product life cycle (PLC), how much volume the product contributes to the total market share of *Apple* (total of the three products *iPhone*, *iPad* and *Mac*). This theme can be further categorised into two subthemes: PLC and market share. PLC is the unit sales curve for a product starting from the time it is launched in the market until it is removed from the market (Rink and Swan 1979). PLC is often used for making strategic functional decision making (Birou et al. 2006). It helps to identify sales opportunities and develop improved forecasts. One of the *iPad* managers considered the PLC of *iPad* while making her/his judgment both in the individual and group rounds. S/he illustrated how *iPad* was losing its position in the market as it had reached the matured stage of PLC. However, s/he had to face a lot of criticism from the other managers regarding this opinion. The second subtheme is the market share of each product, which means what proportion of the total market under the senior manager is represented by each product. Market share is an important entity in any decision making process. *iPhone* is the major player of *Apple* followed by *iPad*. *Mac* contributes the least among these three products to the total market volume. This had an effect on the decision making process. Since *iPhone* contributed the highest, the *iPhone* managers had more say in the group round than other product managers. There were instances noted when the senior managers had more trust on the *iPhone* managers’ judgment than any other managers’ (including himself). This followed with the other managers readjusting their individual forecasts more than the *iPhone* managers, which might consequently result in increased forecast errors. This can be further validated by evaluating the forecast error for all the products, which is covered in section 4.2.

1. Time:

Time pressure plays a major role in the decision making process and emerged as one of the prominent themes during the qualitative data analysis. It takes time to make a decision and the decisions dynamically change with time (Ariely and Zakay 2001). The amount of time allowed for preparation of forecasts is an important factor in selection of a forecasting method (Makridakis and Wheelwright 1989, p. 30). Since the groups were allocated a time limit of 2 hours to finish the experiment (both the individual and group rounds), it had some effect on the decisions made during the group round. The groups took around 30 minutes to complete the individual round where the participants had to make independent adjustments to the SFs. This left them with approximately 90 minutes to reconcile the independent forecasts of all the managers for each quarter. Therefore, the managers had to make quick decisions because of the time constraint. These rushed decisions may not be optimal as they are influenced by intuition and less objectivity. The perception of time pressure does affect the decision quality (Boundless 2015). However, there was one exception to this observation with one group finishing both the rounds in less than 2 hours (~65 minutes). The efforts to achieve consensus during decision making can be quite time-consuming (Harrison 1975, p. 206). Managers do not want to make compromise of some sort; hence, it takes longer to agree on common ground. While generally time pressure is perceived as a barrier to good decision-making, it also helps an organisation to make timely decisions. A limited time frame can help focus the mental energy and effort to make the decisions efficiently. Real or perceived time pressure “may act as a deadline and encourage organisational dexterity” (Boundless 2015).

**4.2 Quantitative data analysis**

Quantitative data is informative and needs to be processed in order to be converted into insights and (depending on the type of research) operationalized suggestions. Different techniques help us to explore, describe and examine relationships and trends in the data (Saunders et al. 2007, pp. 406-407). The participants provided the forecasts for the four quarters of 2015 in the form of numbers. This numerical data is analysed using different descriptive and inferential statistical techniques. Descriptive techniques summarise the data into tables, graphs with the intention of identifying patterns in the data (Teddlie and Tashakkori 2009, pp. 257-258). This helps to understand and interpret the data easily. Inferential techniques are used for testing the descriptive findings. The first and second quarter *Apple* sales reports for 2015 have been released until now (September 2015) and for the purpose of this thesis, the forecasts for only these two quarters are being investigated. The forecasts from the experiment are compared to these actual figures to compute the forecast accuracies. The forecasts for different statistical reconciliation methods (discussed in section 2.3) are generated using the individual forecasts from the first round. The group forecasts are compared with these statistically reconciled forecasts to test whether judgmental reconciliation yields better forecasts than statistical reconciliation. In addition to this, the individual forecasts are compared to the group forecasts to explore the effect of hierarchy and other factors on the reconciliation process.

*4.2.1 Individual versus group forecasts*

Hierarchy plays an important part in group decision making when consensus is to be achieved. During the individual round, the participants make adjustments to the SFs individually and hence independent of any influence from the other group members. However, in the group round all the group members get involved and there can be influences from other members or the other hierarchy that leads to modification of the individual forecasts. The individual forecasts (no-reconciliation method forecasts) without reconciliation are compared to the group reconciled forecasts (table 4.1). It can be seen that the MAPE (average across all groups and both quarters) for the non-reconciled forecasts (24.31) is lower than the group reconciled forecasts (24.67). However, the difference between the MAPEs is not very large. This indicates the presence of some relationship between the individual forecasts and group forecasts. To explore this to a greater extent, deviations from individual forecasts to group forecasts are calculated.

Deviation is defined as

(16)

A positive deviation indicates that the individual forecast was increased during the group reconciliation process and a negative deviation means that the individual forecast was reduced in the group round. The deviation is further calculated in terms of percentages as

(17)

The software package R is used to analyse these deviations for identifying trends among the deviations (Appendix D). The maximum positive deviation is found to be 59.3% corresponding to the senior manager of group G for the first quarter (Q1). The maximum negative deviation is that of group C’s *iPad* manager for second quarter (Q2) amounting to 34.86%. The average deviation across all products, groups and the two quarters Q1 and Q2 is -3.3%. The mean percentage deviation for the upper hierarchy (*Apple*) is 0.42% whereas for the lower hierarchy (*iPhone*, *iPad* and *Mac*), it is a negative average of -4.56%. Moreover, there are 14 cases in total when percentage of deviation is zero, which implies no adjustments were made to the individual forecasts in the group round. Out of these 14 cases, only two of them correspond to the upper hierarchy. For three groups, the *iPad* managers made no adjustment to their individual forecast for the second quarter. More number of negative adjustments (15) were made relative to positive adjustments (11). The boxplots of the percentage deviations categorised for the different products is shown in figure 4.2(a).

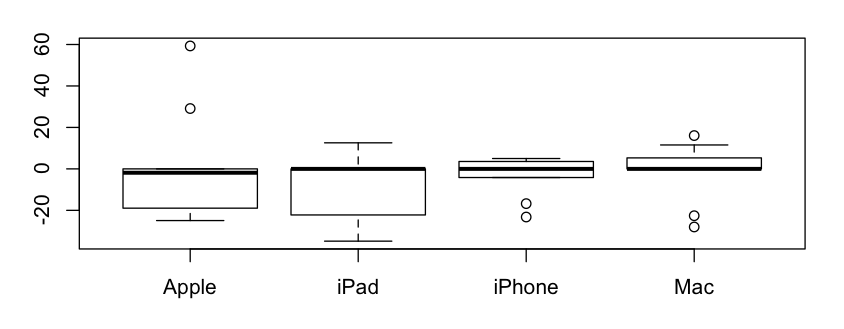


Figure 4.2(a): Boxplot of percentage deviations categorised by product

From figure 4.2 (a), it is observed that all the percentage deviations corresponding to *iPad* are lesser than or equal to zero and those corresponding to *Mac* are greater than or equal to zero. There are two positive outliers with extreme values for the upper hierarchy *Apple*. Another boxplot of percentage deviations categorised by groups is shown in figure 4.2(b).

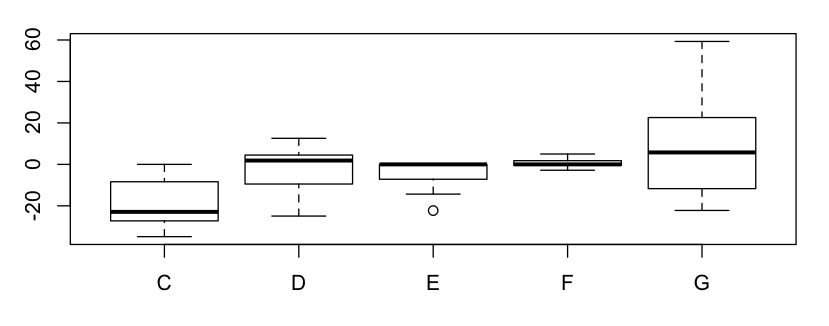


Figure 4.2(b): Boxplot of percentage deviations categorised by group

Group G has the largest spread of adjustments, and the outliers seen in figure 4.2 (a) belong to this group as well. The adjustments of group F are least spread with four out of the ten deviations being zero. The *iPad* and *Mac* managers of this group did not make any adjustment to their individual forecasts for both the first and second quarters. However, group E has the maximum number of zero deviations. For both the quarters, none of the product managers adjusted their individual forecasts and bottom-up reconciliation method was used to obtain the *Apple* forecast. The senior manager’s forecasts were reduced by 14.34% and 22.28% for Q1 and Q2 respectively.

The individual forecasts, group forecasts and the deviations follow non-normal distributions (figure 4.3). Hence, testing the median of deviations against zero using Wilcoxon sign rank test leads to establishing whether the median of the individual forecasts differed significantly from that of the group forecasts. The result is not significant (p= 0.06) at 5% level of significance. Therefore, the individual forecasts do not differ significantly from the group forecasts.

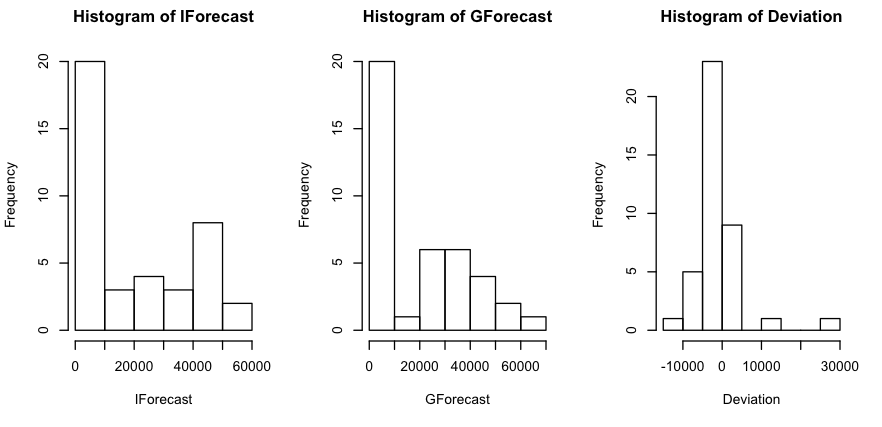


Figure 4.3: Histograms of individual forecasts, group forecasts and deviations

*4.2.2 Judgmental reconciliation versus statistical reconciliation*

The group reconciled forecasts across all the groups can be compared with the forecasts from the standard statistical reconciliation methods: bottom-up, top-down (based on average of historical proportions, proportions of historical averages and forecast proportions) and optimal reconciliation method forecasts. The three top-down methods are termed as top-down1 (average of historical proportions), top-down2 (proportions of historical averages) and top-down3 (forecast proportions) based on their disaggregation technique. The top-down1 and top-down2 proportions are calculated using MS Excel. The individually adjusted forecasts of the first round are used to generate the reconciled forecasts from these statistical reconciliation methods in R (Appendix B). This code calculates the absolute percentage error (APE) and AE (Absolute Error) for each group, product, method and quarter. Based on the results of this code, a table T1 in the format (Group, Product, Quarter, Method, APE and AE) is designed. The MAPEs for the different products are averages of the APEs for all the groups for each time period. To obtain the MAPEs for different products across both the quarters (Q1+Q2) for one method, the APEs of all the groups and both the quarters corresponding to that method are averaged. From table 4.1, the forecast accuracy of the different methods can be compared at the product level. For different quarters, the best reconciliation method for each node/product in terms of forecast improvement are represented in figure 4.4.

Quarter 1

Optimal Reconciliation

Top-Down3

Group

Bottom-Up

Quarter 2

Top-Down

Top-Down2

Group

Group

Quarters 1 & 2

Optimal Reconciliation

Top-Down3

Group

Group

Figure 4.4: MAPE across the groups for different reconciliation methods for quarter 1, 2 and both.

Table 4.1 shows the MAPE across the three dimensions: method, product and time period (Q1, Q2 and Q1+Q2). It can be observed that the top-down3 results in best forecast accuracy across both quarters. However, the difference between the forecast accuracies of the top-down3 and the group reconciliation method is very less. A further evaluation based on relative efficiencies of different reconciliation methods across time series is carried out to check whether group reconciliation method is better than top-down3 and other statistical reconciliation methods.



Table 4.1: Accuracy measurement of different reconciliation methods for the four nodes

The AvgRelMAE is used to evaluate if the judgmentally reconciled forecasts improve/reduce the accuracy compared to the statistical methods. The forecasts from the five statistical reconciliation methods are used as baseline SFs against which the group reconciled forecasts are compared. The table T1 acts as an input for the next R code (Appendix C) that estimates the average MAE for the group reconciled forecasts relative to the statistically reconciled forecasts. There are four time series data corresponding to each node (*Apple*, *iPhone*, *iPad* and *Mac*). The number of non-zero errors of reconciled forecasts for each series is 10 (2 quarters \* 5 groups). Therefore, the total number of available errors for each method for the four time series data is 40. Using these values, the MAEs for all the methods (group reconciliation, bottom-up, top-down1, top-down2, top-down3 and optimal reconciliation) with respect to each time series are calculated. The AvgRelMAE is determined by using the ratio (*ri*) of the MAE of the group reconciliation method to the MAE of any one of the five statistical methods (x) for each time-series (i) as discussed in section 3.3 (equation 14):

The AvgRelMAE of the group reconciliation method when compared with each reconciliation methods is presented in figure 4.5. It can be observed that the AvgRelMAE is lower than one for all the five set of comparisons. This means that on an average the group reconciled forecasts have improved forecast accuracy compared to any other reconciliation method. The AvgRelMAE for group reconciliation relative to top-down3 method is nearly one, which means that on an average both the methods have the same forecast accuracy. This supports the findings from table 4.1.

**Group Reconciliation**

0.9382

0.7328

0.9915

0.8624

0.8032

Figure 4.5: Average Relative MAE of group reconciliation method relative to the other methods

*4.2.3 Deviation and forecast performance*

When the participants were asked the two oral questions after the group round, most of the participants thought their group forecasts were better than the individual forecasts. However, there was not a common answer for the second question on which hierarchy having more influence in the decision making process. Some argued that the upper hierarch had more influence while others thought the product managers had more say in the process. Majority agreed that since both the hierarchies had different inputs (contextual information), it was a collective work where each had some influence on the others. Therefore, the question that arises is if the inter-personal influence have some effect on the forecast accuracy. As described in subsection 4.2.1, this influence is calculated in terms of deviations and deviation percentages between the individual and group round forecasts. The correlation between the forecast deviations and the change in forecast accuracies can help to understand the interplay among these two variables.

Accuracy difference is defined as

(18)

It is the difference between the APE of group forecasts and that of individual forecasts. Accuracy\_difference > 0 means that the group round APE is more than the individual round APE. Therefore, the individual forecasts are better than the group forecasts. Whereas accuracy\_difference < 0 indicates the group reconciled forecasts are better than the individual ones. The difference is further calculated in terms of percentages as

(19)

A table T2 is constructed with the fields – Group, Product, Quarter, Deviation, PDeviation, D\_APE (Accuracy\_difference) and PD\_APE (accuracy\_difference %) for non-zero deviations. The analysis is done in R (Appendix E) and non-parametric tests are considered, as the data is not normally distributed. A correlation (Spearman’s rank) test between Deviation and D\_APE gives an almost negligible coefficient of -0.03. However, as this test is not significant (p=0.88) the relationship between these two variables is further tested at granular hierarchical and group levels. Firstly, the relationship is verified for the two hierarchical levels. For the upper hierarchy, most of the deviations are negative in sign with two positive outliers. However, for any kind of deviation in the upper hierarchy the accuracy reduces during the group round. The forecast error increases whenever the senior manager makes any further adjustment in the group round. The Spearman’s correlation test between the deviations and accuracy\_difference for *Apple* results in a negative correlation of 0.07 but it is not significant (p=0.88). On the other hand, when the lower hierarchy is considered, no such direct relationship can be concluded. 18 instances can be seen when the individual forecasts were readjusted in the group round. Out of these 18 cases, the forecast accuracies improved for 13 of them. The group forecasts for these 13 cases are better than the individual forecasts. The Spearman’s correlation coefficient is 0.18 and it is not significant (p=0.47). Secondly, the effect of deviations on forecast improvement is evaluated at a group level. From figure 4.2(b), the maximum number of deviations are made by group G and the minimum by group F.

For group F, there are two negative deviations that result in positive D\_APE and two positive deviations that result in negative D\_APE. The Spearman’s correlation is almost significant (p=0.05) with a coefficient of -0.95. On the other hand, for group G all the managers have adjusted their independent forecasts except for the *iPhone* manager’s quarter 1 forecast. This results in an negative D\_APE of 2.78 on average. A significant (p=0.02) Spearman’s coefficient ratio of 0.81 is observed for this group. As group E followed the bottom-up method for both the quarters, only the senior manager’s forecasts change. These deviations are negative and lead to decreased forecast accuracy. In case of group C, while the senior manager has made no adjustments, the lower level managers have lowered their forecasts for both the quarters. Half of these have improved the forecasts, whereas the remaining half reduced the accuracy. On an average the D\_APE has a negative value of 4.48. For group D, all the managers make adjustments during the group round for both the quarters excluding the *iPad* manager’s quarter 2 forecast. The Spearman’s correlation tests for group C and D are not significant with coefficients -0.48 and -0.32 respectively. A summary of the Spearman’s rank correlation coefficient between Deviation and D\_APE at two dimensions: hierarchy and group is presented in table 4.2.



Table 4.2: Spearman’s rank correlation coefficient between deviation and D\_APE for different dimensions

The analysis of different types of data collected from the participants was discussed in this chapter. It covered both the qualitative and quantitative data analysis methods. Themes were generated from the study of qualitative data collected in the experiment. On the other hand, different forecast accuracy measures were used to explore the quantitative data. MAPE and AvgRelMAE were calculated for the reconciled forecasts from different statistical aggregation methods and compared with the judgmentally reconciled forecasts. To check the influence of hierarchy on the reconciliation process, deviations of the group forecasts from the individual forecasts were considered. These results have been discussed in line with the two research questions in the next chapter.

**Chapter 5: Discussion**

For the laboratory experiment, a mixed-method approach was employed to collect both qualitative and quantitative data associated with judgmental reconciliation method. Different analysis methods were used to investigate the collected data. The qualitative data was explored using thematic analysis and deviations between the individual and group forecasts were evaluated from the quantitative data. Forecast accuracies were calculated for different reconciliation methods using the measures MAPE and AveRelMAE. The relationship between deviations and forecast performance was elaborated in the last subsection 4.2.3. In this chapter, some of the vignettes from the qualitative data are used to explain the quantitative findings (Bryman 2012, p. 646). The two research questions under study are discussed based on these findings.

**5.1 Research question 1**

*What factors affect the process of judgmentally reconciling hierarchical forecasts?*

The most essential element in group decision making is achieving consensus. This depends on the interaction or communication among the group members (Gilligan et al. 1983 p. 42). In the second round, the group members had to achieve consensus while reconciling their individual independent forecasts. The patterns of behaviour or judgment during the group decision making process can be influenced by a number of factors. The most notable factors (from the thematic analysis) are authority, expertise, product and time. These factors have been identified in the field of business decision making in the literature by a number of authors (Gilligan et al. 1983, p. 42; Goodwin and Wright 2014, p. 319). When different hierarchical levels are involved with the group decision making, the upper hierarchical levels tend to exert major influences on the selection of alternatives (Harrison 1975, p. 200). On the contrary, sometimes the lower level hierarchy having more market knowledge and expertise can guide the upper hierarchy towards better judgments. The differences between the individual and group forecasts also verifies the effect of this influence. From the deviation analysis, it has been observed that the largest adjustments to individual forecasts are made by the upper hierarchical positions. There can be two key possibilities behind these large adjustments. Firstly, the senior management is unaware of most of the product specific information. The senior managers have an overall and general picture of the *Apple* company. The product managers being experts of their products have more detailed knowledge at grass root level about these. Secondly, the product managers have more experience in forecasting and are able to provide suggestions to the senior managers for forecast improvement. “When managers are faced with the task of decision making in the face of uncertainty, their ability to forecast becomes a critical element in the decision-making process”(Hanke and Reitsch 1998, p. 502). The senior manager of group G did not have much experience in forecasting and had to adjust both the Q1 and Q2 values on advice of the product managers during the reconciliation process. Another example is group E, where the managers employed bottom-up approach as the reconciliation method for both the quarters. From the observation notes during the focus group, the senior manager was quoted saying that s/he is willing to adjust her/his forecasts as the product managers were confident about their forecasts. However, these deviations in the *Apple* forecasts reduced the forecast accuracy. The individual forecasts of the senior managers are better than the group reconciled forecasts. On an average, group C has the least MAPE compared to all other groups, and the senior manager of this group has not readjusted her/his individual forecasts in the group round. This proves that authority does help to make better forecasts. Therefore, the senior managers should be more confident about their individual forecasts and make fewer adjustments in the group process.

Market share of each product on the company’s total share drives the course of action during decision making. As *iPhone* has the maximum share amongst the others, *iPhone* managers stick around with their individual forecasts more often compared to the other managers. The *iPad* and *Mac* managers had to make further adjustments to their individual forecasts. However, adjustments made by the lower hierarchy in the group round usually results in decreased forecast errors. As *iPad* and *Mac* managers have made more adjustments relative to *iPhone* managers, their forecast performances are better than the *iPhone* managers. Therefore, the managers need to trust each other irrespective of their products’ market shares in order to accept recommendations on further adjustments. Time is another vital attribute for effective group decision making. The groups have to agree on a set of reconciled forecasts within the allocated time period. To do so, they have to discuss the rationale behind the individual forecasts and readjust them based on some new information, which takes a substantial amount of time. There are changes in the thought process of managers and it takes time to process information and make decisions. However, the reconciled forecasts of the group F (that took the least time to complete the task) are better than any other group, except group C. The groups (D, E and G) overthought about the forecasts as there was time available and it affected the forecast quality. The members of group F stayed focussed that helped them to render adequate decisions in less time. Nevertheless, managers should not be too quick to jump into conclusions as it can distort the decision making process.

**5.2 Research question 2**

*Does judgmental reconciliation result in better forecast accuracy than statistical reconciliation methods?*

Unlike mathematical aggregation, behavioural aggregation is a method of reconciling the forecasts through group consensus. It considers an unequal weighting of individual forecasts to form consensus (Sniezek 1990). None of the groups has employed any of the statistical reconciliation methods as the standard reconciliation method across all the four quarters of 2015. The most plausible reason behind this is every statistical method has some drawback of aggregating the forecasts. If the bottom-up method is applied, judgmental adjustments of the top hierarchy are not reflected in the reconciled forecast. Thus, the senior managers find it difficult to accept these reconciled forecasts. Only the senior manager of group E accepted the bottom-up forecasts for the first two quarters, as s/he was less confident for her/his judgment than the product managers. The three top-down reconciliation methods, on the contrary, neglect the information related to the lower level products. Hence, the product managers of the lower hierarchy are not satisfied with the disaggregated forecasts. Being experts in this field, the product managers know better about their products than the statistical tools and hence they are not willing to give up their opinions regarding the forecasts. Even when the top-down methods are used for reconciliation, the groups decide on the disaggregation proportions according to the available contextual information. Although optimal reconciliation has been found to outperform all other reconciliation methods in the literature (Hyndman et al. 2011), none of the groups has considered it for reconciling forecasts of any quarter. Either they are not aware of this method or it is too complicated to be employed without the help of statistical tools. All the groups (except for group E) have adopted the judgmental reconciliation method for reconciliation in the group round. This method has the advantage of incorporating ad hoc adjustments into individual forecasts belonging to all hierarchical nodes in light of new information. This is very common in the practical world as all the managers do not have access to the company information (like the upper hierarchy) or about some specific products (like the lower hierarchy) that might affect the sales of their products. Like for example, in this experiment the *Mac* managers have no knowledge about the launch of 13” *iPad* unless the respective *iPad* managers discuss it in the group round. The managers have considered the fact that the 13” *iPad* will affect the *Mac* sales, and readjusted the *Mac* sales figures accordingly during reconciliation. Another example can be of the upper hierarchy having extra information. The senior managers have news about a collaboration deal between *Apple* and *Microsoft* that would further affect the sales of the three products: *iPhone*, *iPad* and *Mac*. The participants have quoted this deal as one reason why they are more confident about their group forecasts than their individual forecasts, during the second oral question. Another characteristic of this method is all the managers collectively decide the weights assigned to different forecasts during reconciliation. Unlike statistical reconciliation methods, these weights do not depend on any historical data or observed data. Therefore, the statistical reconciliation methods cannot incorporate the contextual information in an optimal manner. From the MAPE analysis in the previous chapter (table 4.1), it can be seen that the best reconciliation methods are the top down method (based on forecasted proportions) followed by the group reconciliation method. However, the AveRelMAE analysis shows that on an average the group reconciliation method has improved forecast accuracy than any other reconciliation method. At a product level, the group reconciliation method is best for *iPad* and *Mac* for both quarters (Q1 and Q2) of 2015.

On the other hand, a number of interrelated elements are associated with the judgmental reconciliation method. Human bias is one of the most important element. Most of the managers have biased opinions about the forecasts and while revisiting their rationale for their judgments during the group round they put more weightage to their own opinion than the others’. This confirms what Soll and Mannes (2011) have concluded from their experiment in the past. The managers from two groups have agreed that during the individual round they make insufficient forecast adjustments as their judgments are anchored to a particular set of forecast values (SFs). The group discussions help them to reduce this bias and make better forecasts. For example, the *iPad* managers are very optimistic while adjusting the SFs during the individual round. They take note of this optimistic behaviour during the group round and readjust their forecasts, resulting in a large number of negative deviations in the group round (figure 4.4 (a)). However, these negative forecasts have reduced the forecast errors drastically for the group forecasts. The bottom up and top down are the best reconciliation method for *Mac* (quarter 1) and *Apple* (quarter 2) according to the MAPE statistics in table 4.1. If the respective managers had not readjusted their individual forecasts, the forecast error would have been reduced. Therefore, along with the advantages of the judgmental reconciliation method there are a number of limitations as well.

This chapter discussed the two research questions in light of the analysed findings from the previous chapter. There are a number of factors that affect the judgmental reconciliation process. Both the hierarchies have influence on the decisions being made. The upper hierarchy has the leadership impact while the lower hierarchy has the advantage of having more expertise knowledge. Adjustments made to the independent forecasts of the upper hierarchy in the group round decreases the forecast accuracy. On the contrary, it acts as a boon for the lower hierarchy. Judgmental reconciliation through group consensus is found to be the best aggregation method for hierarchical forecasts. However, it has the disadvantage of being biased and vulnerable to peer pressure.

**Chapter 6: Summary and conclusion**

**6.1 Summary**

When judgmental adjustments are performed at different echelons of a hierarchical structure, a group reconciliation method is necessary to reconcile these adjusted forecasts. It is an ‘interpersonal reconciliation process’ to nurture consensus among managers from various hierarchies, which makes them interdependent on each other (Syntetos et al. 2015). A behavioural laboratory experiment was conducted to see how consensus is achieved while judgmentally reconciling the hierarchical forecasts. The participants were divided into groups of four with each one representing a node of a 1-3 business pyramid. They were provided with the quarterly statistical forecasts for 2015 and some contextual information about the future sales. They were asked to adjust the SFs individually in the first round and then in the second round judgmentally reconcile these forecasts with the other group members such that the final forecasts add up the hierarchy. This process of making future estimates involves a rational process of “extending historical data and experience into the future” (Hanke and Reitsch 1998, p. 507). The group round data was evaluated and themes based on the factors that affect the decision making process were identified. The individual forecasts were used as baseline forecasts to produce reconciled forecasts using different statistical reconciliation methods. Accuracies of different types of forecasts were calculated using two accuracy measures: MAPE and AvgRelMAE. A comparison of the forecast errors between the group forecasts and different sets of the statistically reconciled forecasts was drawn. Group forecasts were found to the best forecasts as it helped in incorporation of additional managerial information at all hierarchical levels. Furthermore, deviations between the forecasts of the two rounds: individual and groups were computed and insights were drawn as to how these deviations might be related to the qualitative themes or the forecast performance. Based on these findings, implications for the academic circle and practitioners are presented in section 6.2. The contributions and limitations of this study are explained in the subsequent sections.

**6.2 Implications for theorists and practitioners of forecasting**

The theoretical and practical implications from the findings of the laboratory experiment are considered in this section. Group decision making processes dominate the business world with senior managers spending more than 80% of each working day in one sort of group or another (Gilligan et al. 1983, p. 42). The most important characteristic of these group processes is the element of attaining consensus for a given purpose. Since the managers are in personal contact with one another, the weight allocated to each manager’s judgment depends in large part on the role and personality of that manager in the organisation (Makridakis and Wheelwright 1989, p. 241). Therefore, for effective decision making it is important to understand the anatomy of the group. Additionally, investments in judgmental forecasting can help reap considerable financial returns (Syntetos et al. 2010). The new insights implied by this study regarding judgmental reconciliation of hierarchical forecasts are:

1. The best method for reconciliation of hierarchical forecasts is through judgmental reconciliation. In this method, judgmental interventions are allowed at all levels and hence no contextual data is lost in the process. Additionally, group discussions help to reduce biases and make the managers more proficient because of their greater amount of knowledge and experience. The lower hierarchy tends to benefit more in the group decision making process than the upper hierarchy. Therefore, the managers from different hierarchical levels should work in harmony and decide on a set of common forecasts.
2. As judgmental reconciliation is a collective approach to decision making, there are a number of factors that influence the patterns of behaviour. Upper hierarchy has the leadership influence on the lower hierarchies and the lower hierarchy managers have the effect of being experts in their respective fields. The authority effect of the senior management facilitates in making better decisions. On the contrary, when the upper hierarchy make judgmental adjustments because of pressure from the product managers it results in decreased forecast accuracy. When minimum adjustments are required to reconcile the forecasts, positive adjustments (increasing the individual forecasts) improves the forecast performance whereas negative adjustments (reducing the individual forecasts) deteriorates the forecast accuracy. Thus, managers should be careful while arriving at any decision regarding the forecasts to avoid any pressure from other group members.

**6.3 Research contributions**

The key contribution of this study has been in terms of research methodology. It is the first time a laboratory experiment with mixed method approach has been conducted in the field of judgmental forecasting. In behavioural aggregation, the group members decide on a group judgment by communicating in an open discussion or through a more structured communication process (Goodwin and Wright 2014, p. 307). Therefore, both qualitative and quantitative data are needed to get better insights of how decision makers would handle uncertainty while exploiting their experiences in the forecasting domain. Moreover, parallel mixed data strategies comprising of thematic (qualitative) and quantitative analysis methods are used to examine the collected data. Another major contribution is the establishment of group reconciliation method as the best aggregation method when judgmental interventions are performed at various levels of a hierarchy. This method is an interpersonal reconciliation process that encourages participative managerial skills where one does not get to take full ownership of the forecasts (Gilligan et al. 1983, p. 35). This addresses a significant research area where further experimentation and theoretical research has been called for by many authors (Petropoulos 2015, Spithourakis et al. 2015, Syntetos et al. 2015a). As a result of this work, a number of theoretical and practical implications have been recognised.

**6.4 Limitations and future directions**

Despite the major contributions of this study, it leaves some open questions as well. First, the research sample included only five focus groups (20 participants) that resulted in insufficient data to analyse. The credibility of the derived conclusions is not very high. Like the results show that time taken to complete the task has some effect on the forecast accuracy. However, only one group’s data was available to make this deduction since only one group had finished the task before the allocated time. Future research is needed to conduct such an experiment with more number of focus groups so that the dependability of the results become stronger. Additionally, only two-step ahead forecasts (quarter 1 and 2) were analysed for the purpose of this study. When sales data for all the quarters of 2015 are accessible, the analysis should be repeated for all the four quarters to determine if the conclusions still hold true. Therefore, there arises the question whether the results can be generalised to a larger population. Second, this experiment did not explore the many ways in which the data availability can vary and how these variations can shape the judgmental reconciliation process. The participants were provided with qualitative information about what is expected from *Apple* or the three products in 2015. If they are also provided with the contextual information about past special events that had an effect on the sales, it might have some influence in the aggregation method. Graphs were used to display the historical time series data along with the SFs in the questionnaires. Past experiments have confirmed the effects of data presentation format on judgmental forecasting (Harvey and Bolger 1996). Therefore, experiments should be designed and analysed with data being displayed in different formats like tabular form or graphical form or both, to understand the underlying cognitive processes (Harvey and Bolger 1996). The fact that case company was not kept anonymous to the participants might have introduced some bias (depending on whether they liked or disliked Apple products) in the results. Another version of this experiment can be designed to evaluate this bias by not letting the participants know about the case company during the task. Moreover, participants had to make multiple decisions within a short time span (2 hours). Whereas if they are allowed to deliberate extensively over each forecast without any time limit, it might have an influence on their performance. Therefore, a further objective of the current experimental design would be to see how variations in data availability and time allocation could bring changes in the decisions specified. Third, as many researchers consider students unrepresentative of the real world decision makers, relevant participants (managers) can be recruited to join the experiment. Real practitioners may have acquired expert skills in forecasting because of regular experience of carrying out the task (Goodwin and Wright 2014, pp. 262-263). The way they tackle judgmental tasks in the real world may be different from the way the students approach them. Additional studies are required to confirm if changes in the nature of a judgmental task can have any major changes in the way the problem is handled. Empirical research that extends this unchartered territory of judgmental reconciliation of various levels of hierarchical forecasts is merited.

**Bibliography:**

* Anderson, C. & Brown, C. E. 2010. The functions and dysfunctions of hierarchy. *Research in Organizational Behavior* 30 (1), pp. 55–89.
* Apple. 2015. *Apple info* [Online] Available at: <http://www.apple.com/about/> [Assessed: 21 February 2015].
* Ariely, D. and Zakay, D. 2001. A timely account of duration in decision making. *Acta Psychologica* 108 (1), pp. 187-207.
* Athanasopoulos, G., Ahmed, R. A. and Hyndman, R. J. 2009. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting* 25 (1), pp. 146-166.
* Beckenstein, A. R. 1987. Forecasting and the environment: the challenges of rapid change. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 279-291.
* Bello, D., Leung, K., Radebaugh, L., Tung, R. L. and Witteloostuijn, A. V. 2009. From the editors: student samples in international business research*. Journal of International Business Studies* 40 (1), pp. 361–364.
* Birou, L. M., Fawcett, S. E. and Magnan, G. M. 2006. The Product Life Cycle: A Tool for Functional Strategic Alignment. *International Journal of Purchasing and Materials Management* 34 (1), pp. 37-52.
* Blattberg, R. C. and Hoch, S. J. 1990. Database models and managerial intuition: 50% model + 50% manager. *Management Science* 36 (8), pp. 887-899.
* Blumber, B. F., Cooper, D. R. and Schindler, P. S. 2014. *Business research methods*. 4th Ed. London: McGraw-Hill Higher Education.
* Boundless. 2015. *Time Pressure as a Barrier to Decision Making*. Boundless Management [Online]. Available at: <https://www.boundless.com/management/textbooks/boundless-management-textbook/decision-making-10/barriers-to-decision-making-80/time-pressure-as-a-barrier-to-decision-making-386-585/> [Accessed: 13 August 2015].
* Bryman, A. 2012. *Social research methods*. 4th Ed. Oxford: Oxford University Press.
* Bryman, A. and Bell, E. 2007. *Business research methods.* 2nd Ed. Oxford: Oxford University Press.
* Cohen, M. S., Freeman, J. T. and Wolf, S. 1996. Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38 (2), pp. 206-219.
* Collis, J. and Hussey, R. 2009. *Business research: a practical guide for undergraduate & postgraduate students*. 3rd Ed. England: Palgrave Macmillan.
* Cooper, D. R. and Schindler, P. S. 2006. *Business Research Methods*. 9th Ed. Boston: McGraw-Hill.
* Dangerfield, B. J. and Morris, J. S. 1992. Top-down or bottom-up: aggregate versus disaggregate extrapolations. *International Journal of Forecasting* 8 (2), pp. 233-241.
* Deck, C. and Smith, V. 2013. Using laboratory experiments in logistics and supply chain research. *Journal of Business Logistics* 34 (1), pp. 6-14.
* Diamantopoulos, A. and Mathews, B. P. 1989. Factors affecting the nature and effectiveness of subjective revision in sales forecasting: An empirical study. *Managerial and Decision Economics* 10(1), pp. 51–59.
* Easterby-Smith, M., Thorpe, R. and Jackson, P. 2012. *Management Research*. 4th Ed. London: Sage.
* Edmundson, R., Lawrence, M. and O’Connor, M. 1988. The use of non time series information in sales forecasting. *Journal of Forecasting* 7 (3), pp. 201-211.
* Fildes, R. 1987. Forecasting: the issues. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 150-172.
* Fildes, R., Goodwin, P. and Onkal, D. 2015. *Information use in supply chain forecasting*. Lancaster University Department of Management Science working paper series, no. 2, vol. 2015, Department of Management Science, Lancaster University, Lancaster.
* Fildes, R., Goodwin, P., Lawrence, M. and Nikolopoulos, K. 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* 25 (1), pp. 3-23.
* Firth, M. 1977. *Forecasting methods in business and management*. London: Edward Arnold.
* Fliedner, G. 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers and Operations Research* 26 (10-11), pp. 1133-1149.
* Fliedner, G. 2001. Hierarchical forecasting: issues and use guidelines. *Industrial Management & Data Systems* 101 (1), pp. 5 – 12.
* Franses, P. H. and Legerstee, R. 2009. Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting* 25 (1), pp. 35-47.
* Franses, P. H. and Legerstee, R. 2010. Do experts’ adjustments on model-based SKU-level forecasts improve forecast accuracy. *Journal of Forecasting* 29 (3), pp. 331–340.
* Franses, P. H. and Legerstee, R. 2011. Experts’ adjustment to model-based SKU-level forecasts: does the forecast horizon matter? *Journal of Operational Research Society* 62 (1), pp. 537-543.
* Franses, P. H. and Legerstee, R. 2013. Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting* 29 (1), pp. 80-87.
* Franses, P. H., Kranendonk, H. C. and Lanser, D. 2011. One model and various experts: evaluating Dutch macroeconomic forecasts. *International Journal of Forecasting* 27 (1), pp. 482-495.
* Franses, P. H., McAleer, M. and Legerstee, R. 2009. Expert opinion versus expertise in forecasting. *Statistica Neerlandica* 63 (3), pp. 334-346.
* Ghauri, P. and Grønhaug, K. 2002. *Research methods in business studies: a practical guide*. 2nd Ed. England: Pearson Education Limited.
* Gilligan, C., Neale, B. and Murray, D. 1983*. Business decision making*. Oxford: Philip Allan.
* Goodwin, P. 2000. Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting* 16 (1), pp. 261-275.
* Goodwin, P. 2002. Integrating management judgment with statistical methods to improve short-term forecasts. *Omega- International Journal of Management Science* 30 (2), pp. 127-135.
* Goodwin, P. and Fildes, R. 1999. Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioural Decision Making* 12 (1), pp. 37-53.
* Goodwin, P and Wright, G. 2014. *Decision analysis for management judgment*. 5th Ed. Cornwall: Wiley.
* Graefe, A. and Armstrong, J. S. 2011. Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting* 27 (1), pp. 183-195.
* Gross, C. W. and Peterson, R. T. 1983. *Business forecasting*. 2nd Ed. Boston: Houghton Mifflin Company.
* Guba, E. G. and Lincoln, Y. S. 2005. Paradigmatic controversies, contradictions and emerging confluences. In Denzin, N. and Lincoln, Y. S. (Eds.) *The sage handbook of qualitative research*. 3rd Ed. Thousand Oaks: Sage Publications, pp. 191-215.
* Gupta, U. G. and Clarke, R. E. 1996. Theory and applications of the Delphi technique: a bibliography (1975-1994). *Technological Forecasting and Social Changes* 53 (2), pp. 185-211.
* Hanke, J. E. and Reitsch, A. G. 1998. *Business forecasting*. 6th Ed. Upper Saddle River, NJ: Prentice-Hall.
* Hannafin, S. 2004. *The Delphi technique: A methodology to support the development of a national set of child well-being indicators.* The National Children’s Office, Dublin.
* Harrison, E. F. 1975. *The managerial decision-making process*. Boston: Houghton Mifflin Company.
* Harvey, N. 1995. Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes* 63 (3), pp. 2476-263.
* Harvey, N. and Bolger, F. 1996. Graphs versus tables: effects of data presentation format on judgemental forecasting. *International Journal of Forecasting* 12 (1), pp. 119-137.
* Hibon, M. and Evgeniou, T. 2005. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting* 21 (1), pp. 15-24.
* Highhouse, S. and Gillespie, J. Z. 2010. Do samples really matter that much. In Lance, C. E. and Vandenberg, R. J. (Eds.) *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences*. Routledge: Taylor and Francis group, pp. 247-265.
* Hogarth, R. M. 1978. A note on aggregating opinions. *Organizational Behavior and Human Performance* 21 (1), pp. 40-46.
* Hogarth, R. M. and Makridakis, S. 1981. Forecasting and planning: an evaluation. *Management Science* 27 (2), pp. 115-138.
* Hyndman, R. J. and Athanasopoulos, G. 2013a. *Forecasting: principles and practice*. Section 3/1. [Online]. Available at: https://www.otexts.org/fpp/3/1 [Accessed on: 23 August 2015].
* Hyndman, R. J. and Athanasopoulos, G. 2013b*. Forecasting: principles and practice*. Section 9/4. [Online]. Available at: https://www.otexts.org/fpp/9/4 [Accessed on: 4 August 2015].
* Hyndman, R. J. and Athanasopoulos, G. 2014. Optimally reconciling forecasts in a hierarchy. *Foresight* 35 (Fall 2014), pp. 42-48.
* Hyndman, R. J. and Koehler, A. B. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22 (1), pp. 679-688.
* Hyndman, R. J., Ahmed, R. A. and Athanasopoulos, G. 2007. *Optimal combination forecasts for hierarchical time series*. Working paper 09/07. Department of Econometrics and Business Statistics, Monash University. [Online]. Available at: http://webdoc.sub.gwdg.de/ebook/serien/e/monash\_univ/wp9-07.pdf [Accessed: 5 August 2015]
* Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. 2011. Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis* 55 (9), pp. 2579- 2589.
* Kinney, W. R., Jr. 1971. Predicting earnings: Entity versus subentity data. *Journal of Accounting Research* 9 (1), pp. 127-136.
* Krueger, R. A. and Casey, M. A. 2015. *Focus groups: a practical guide for applied research*. 5th Ed. London: Sage.
* Landeta, J. and Barrutia, J. 2011. People consultation to construct the future: a Delphi application. *International Journal of Forecasting* 27 (1), pp. 134-151.
* Lawrence, M., Goodwin, P., O’Connor, M. and Önkal, D. 2006. Judgmental forecasting: a review of progress over the last 25 years. *International Journal of Forecasting* 22 (1), pp. 493-518.
* Legerstee, R. and Franses, P. H. 2014. Do experts’ SKU forecasts improve after feedback? *Journal of Forecasting* 33 (1), pp. 69-79.
* Leonard, M. 2014. Forecasting short seasonal time series using aggregate and analogous series. *Foresight guidebook*, pp. 26-29.
* Lim, J. S. and O'Connor, M. 1995. Judgemental adjustment of initial forecasts: Its effectiveness and biases. *Journal of Behavioral Decision Making* 8 (3), pp. 149-168.
* Lincoln, Y. S. and Guba, E. G. 1985. *Naturalistic Inquiry*. London: Sage.
* Mahmoud, E. 1987. The evaluation of forecasts. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 508-522.
* Makridakis, S. and Wheelwright, S. C. 1977. Forecasting: issues and challenges for marketing management. *Journal of Marketing* 41 (4), pp. 24-38.
* Makridakis, S. and Wheelwright, S. C. 1987. Introduction to management forecasting. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 248-265.
* Makridakis, S. and Wheelwright, S. C. 1989. *Forecasting methods for management*. 5th Ed. United States of America: John Wiley & Sons.
* Mathews, B. P. and Diamantopoulos, A. 1986. Managerial intervention in forecasting: An empirical investigation of forecast manipulation. *International Journal of Research in Marketing* 3 (1), pp. 3–10.
* Mathews, B. P. and Diamantopoulos, A. 1989. Judgemental revision of sales forecasts- a longitudinal extension. *Journal of Forecasting* 8 (1), pp. 129–140.
* Mathews, B. P. and Diamantopoulos, A. 1990. Judgmental revision of sales forecasts- effectiveness of forecast selection. *Journal of Forecasting* 9 (1), pp. 407–415.
* Mathews, B. P. and Diamantopoulos, A. 1992. Judgemental revision of sales forecasts- the relative performance of judgementally revised versus non revised forecasts. *Journal of Forecasting* 11 (1), pp. 569–576.
* Mathews, B. P. and Diamantopoulos, A. 1994. Towards a taxonomy of forecast error measures- a factor-comparative investigation of forecast error dimensions. *Journal of Forecasting* 13 (1), pp. 409–416.
* May, T. 2001. *Social Research: Issues, methods and process*. 3rd Ed. Buckingham: Open University Press.
* Ord, K. and Fildes, R. 2013. *Principles of business forecasting*. International Ed. South-Western: Cengage Learning.
* Petropoulos, F. 2014*. Judgmental reconciliation of hierarchical forecasts*. [Online]. 25 June 2014. Available at: http://www.forsoc.net/2014/06/25/judgmental-reconciliation-of-hierarchical-forecasts/ [Accessed: 5 August 2015].
* Petropoulos, F. 2015. Forecasting support systems: ways forward. *Foresight* [In press].
* Petropoulos, F., Fildes, R. and Goodwin, P. 2015. Do ‘big losses’ in judgmental adjustments to SFs affect experts’ behaviour? *European Journal of Operations Research* [In press].
* Rink, D. R. and Swan, J. E. 1979. Product life cycle research: A literature review. *Journal of Business Research* 7 (3), pp. 219-242.
* Robson, C. 2002. *Real World Research: a research for social scientists and practitioner-researchers*. 2nd Ed. Oxford: Blackwell Publishing.
* Saunders, M., Lewis, P. and Thornhill. 2007. *Research methods for business students*. 4th Ed. England: Pearson Education Limited.
* Scharzkopf, A. B., Tersine, R. J. and Morris, J. S. 1988. Top-down versus bottom-up forecasting strategies. *International Journal of Production Research* 26 (11), pp. 1833-1843.
* Siemsen, E. 2011. The usefulness of behavioral laboratory experiments in supply chain management research. *Journal of Supply Chain Management* 47 (3), pp. 17–18.
* Sniezek, J. A. 1990. A comparison of techniques for judgmental forecasting by groups with common techniques. *Group & Organization Studies* 15 (1), pp. 5-20.
* Soll, J. B. and Mannes, A. E. 2011. Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting* 27 (1), pp. 81-102.
* Spithourakis, G., Petropoulos, F., Nikolopoulos, K. and Assimakopoulos, V. 2015. Amplifying the learning effects via a forecasting and foresight support system*. International Journal of Forecasting* 31 (1), pp. 20-32.
* Stewart, T. R. 2001. Improving reliability of judgmental forecasting, In Armstrong, J. S. (Ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. New York: Kluwer Academic Publishers, pp. 81-106.
* Syntetos, A.A., Babai, Z., Boylan, J., Kolassa, S. and Nikolopoulos, K. 2015a. Supply chain forecasting: theory, practice, their gap and the future*. European Journal of Operations Research*. [Under review].
* Syntetos, A. A., Georgantzas, N. C., Boylan, J. E. and Dangerfield, J. E. 2011. gloent and supply chain dynamics. *Journal of the Operational Research Society* 62 (1), pp. 1138–1158.
* Syntetos, A.A., Kholidasari, I. and Naim, M. M. 2015b. The effects of integrating management judgment into stock replenishment decisions. *European Journal of Operations Research*. [Under review].
* Syntetos, A. A., Konstantinos, N. and Boylan, J. E. 2010. Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *International journal of forecasting* 26 (1), pp. 134-143.
* Syntetos, A. A., Konstantinos, N., Boylan, J. E, Fildes, R. and Goodwin, P. 2009. The effects of integrating management judgment into intermittent demand forecasts*. International Journal of Production Economics* 118 (1), pp. 72-81.
* Teddlie, C. and Tashakkori, A. 2009. *Foundations of mixed method research*. London: Sage.
* Theil, H. 1954. *Linear aggregation of economic relations*. Amsterdam: North-Holland Publishing Company.
* Thomas, R. W. 2011. When student samples make sense in logistics research. *Journal of Business Logistics* 32 (3), pp. 287–290.
* Thomas, R.W., Esper, T.L. and Stank, T.P. 2010. Testing the Negative Eﬀects of Time Pressure in Retail Supply Chain Relationships. *Journal of Retailing* 86 (4), pp. 368–382.
* Trapero, J. R., Pedregal, D. J., Fildes, R. and Kourentzes, N. 2013. Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting* 29 (2), pp.234-243.
* Weinstein, D. 1987. Forecasting for industrial products. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 450-467.
* Winkler, R. L. 1987. Judgmental and bayesian forecasting. In Makridakis, S. and Wheelwright, S. C. (Eds.) *The handbook of forecasting*. 2nd Ed. United States of America: John Wiley & Sons, pp. 248-265.
* Zellner, A. and Tobias, J. 2000. A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting* 19 (5), pp. 457-469.
* Zotteri, G., Kalchschmidt, M. and Saccani, N. (2014). Forecasting by cross-sectional aggregation. *Foresight* 35 (Fall 2014), pp. 35-41.

**Appendix A**

**R code for statistical forecasts for questionnaires**

**iPhone (AAN):**

a=read.table("clipboard",header=T)

a

insample = ts(a, start=c(2011+3/4), frequency=4)

insample

x = c("2012-Q1", "2013-Q1", "2014-Q1", "2015-Q1")

y = as.vector(insample)

plot(y, type="o", xaxt = "n", xlim=c(1,18), ylim=c(10000, 70000),

xlab="Time (quarters)", ylab="iPhone Sales (in millions $)")

axis(1, at=c(2,6,10,14), labels=x)

abline(v=13.5, lty=2, col="lightgray")

library(Mcomp)

fit=ets(insample, model="AAN", damped=TRUE)

summary(fit)

lines(14:17, forecast(fit,h=4)$mean, lty=2, type="o", col="blue")

b=round(forecast(fit,h=4)$mean,digits=2)

legend(14,70000,c("Actuals","Forecast"), lty=c(1,2),lwd=c(1,1), col=c("black", "blue") )

text( 14.5, b[1]+1800, b[1], pos=2, cex= 0.8, col ="red")

text( 15, b[2]-500, b[2], pos=1, cex= 0.8, col ="red")

text( 16, b[3]+500, b[3], pos=3, cex= 0.8, col ="red")

text( 16.5, b[4]-1800, b[4], pos=4, cex= 0.8, col ="red")

**iPad (ANN):**

a=read.table("clipboard",header=T)

a

insample = ts(a, start=c(2011+3/4), frequency=4)

insample

x = c("2012-Q1", "2013-Q1", "2014-Q1", "2015-Q1")

y = as.vector(insample)

plot(y, type="o", xaxt = "n", xlim=c(1,18), ylim=c(3000, 15000),

xlab="Time (quarters)", ylab="iPad Sales (in millions $)")

axis(1, at=c(2,6,10,14), labels=x)

abline(v=13.5, lty=2, col="lightgray")

library(Mcomp)

fit=ets(insample, model="ANN", damped=FALSE)

summary(fit)

lines(14:17, forecast(fit,h=4)$mean, lty=2, type="o", col="blue")

b=round(forecast(fit,h=4)$mean,digits=2)

legend(14,14000,c("Actuals","Forecast"), lty=c(1,2),lwd=c(1,1), col=c("black", "blue") )

text( 15, b[1]+500, b[1], pos=2, cex= 0.8, col ="red")

text( 15, b[2]-200, b[2], pos=1, cex= 0.8, col ="red")

text( 16, b[3]+200, b[3], pos=3, cex= 0.8, col ="red")

text( 16.5, b[4]-500, b[4], pos=4, cex= 0.8, col ="red")

**MAC (ANN):**

a=read.table("clipboard",header=T)

a

insample = ts(a, start=c(2011+3/4), frequency=4)

insample

x = c("2012-Q1", "2013-Q1", "2014-Q1", "2015-Q1")

y = as.vector(insample)

plot(y, type="o", xaxt = "n", xlim=c(1,18), ylim=c(4000, 8000),

xlab="Time (quarters)", ylab="MAC Sales (in millions $)")

axis(1, at=c(2,6,10,14), labels=x)

abline(v=13.5, lty=2, col="lightgray")

library(Mcomp)

fit=ets(insample, model="ANN", damped=FALSE)

summary(fit)

lines(14:17, forecast(fit,h=4)$mean, lty=2, type="o", col="blue")

b=round(forecast(fit,h=4)$mean,digits=2)

legend(14,8000,c("Actuals","Forecast"), lty=c(1,2),lwd=c(1,1), col=c("black", "blue") )

text( 15, b[1]+100, b[1], pos=2, cex= 0.8, col ="red")

text( 15, b[2]-50, b[2], pos=1, cex= 0.8, col ="red")

text( 16, b[3]+50, b[3], pos=3, cex= 0.8, col ="red")

text( 16.5, b[4]-100, b[4], pos=4, cex= 0.8, col ="red")

**Apple (AAN):**

a=read.table("clipboard",header=T)

a

insample = ts(a, start=c(2011+3/4), frequency=4)

insample

x = c("2012-Q1", "2013-Q1", "2014-Q1", "2015-Q1")

y = as.vector(insample)

plot(y, type="o", xaxt = "n", xlim=c(1,18), ylim=c(20000, 75000),

xlab="Time (quarters)", ylab="Total Sales (in millions $)")

axis(1, at=c(2,6,10,14), labels=x)

abline(v=13.5, lty=2, col="lightgray")

library(Mcomp)

fit=ets(insample, model="AAN", damped=TRUE)

summary(fit)

lines(14:17, forecast(fit,h=4)$mean, lty=2, type="o", col="blue")

b=round(forecast(fit,h=4)$mean,digits=2)

legend(14,75000,c("Actuals","Forecast"), lty=c(1,2),lwd=c(1,1), col=c("black", "blue") )

text( 14.5, b[1]+1500, b[1], pos=2, cex= 0.8, col ="red")

text( 15, b[2]-500, b[2], pos=1, cex= 0.8, col ="red")

text( 16, b[3]+500, b[3], pos=3, cex= 0.8, col ="red")

text( 16.5, b[4]-1500, b[4], pos=4, cex= 0.8, col ="red")

**Appendix B**

**R code for statistical aggregation methods and error measures**

data=read.table(pipe("pbpaste"),header=T)

head(data)

attach(data)

#bottom level nodes (bln)

bln = 3

#define S matrix

S = matrix(c(1,1,1,1,0,0,0,1,0,0,0,1), nrow=bln+1, ncol=bln, byrow=TRUE)

S

#define P table for bottom-up (Athanasopoulos et al, eq.3)

PBU = matrix(0, nrow=bln, ncol=1+bln)

PBU[,(2):(1+bln)] = diag(bln)

PBU

#define P table for top-down, based on average historical proportions

PTD1=c(0.6337, 0.2087, 0.1576)

FTD1= c(fcs[1], PTD1\*fcs[1])

FTD1

#define P table for top-down, based on proportions of historical averages

PTD2=c(0.6449, 0.2040, 0.1511)

#define P table for top-down, based on forecast proportions (Athanasopoulos et al, eq.7)

PTD = matrix(0, nrow=bln, ncol=1+bln)

#define P table for optimal combination (Athanasopoulos et al, eq.10)

library(MASS) #for ginv() function

POP = ginv(t(S) %\*% S) %\*% t(S)

POP

for (i in 1:5){

paste("Group",i)

actuals=Actual[Quarter=="Q1" & Group==i]

fcs=IForecast[Quarter=="Q1" & Group==i]

gfcs=GForecast[Quarter=="Q1" & Group==i]

#the reconciled forecasts for each approach

FBU = S %\*% PBU %\*% fcs

FTD1= c(fcs[1], PTD1\*fcs[1])

FTD2= c(fcs[1], PTD2\*fcs[1])

PTD[,1] = fcs[(2):(1+bln)]/sum(fcs[(2):(1+bln)])

FTD3 = S %\*% PTD %\*% fcs

FOP = S %\*% POP %\*% fcs

print(fcs)

print(FBU)

print(FTD1)

print(FTD2)

print(FTD3)

print(FOP)

print(gfcs)

#calculate the Absolute Error for each reconciliation approach

AE\_BU = abs(actuals - FBU)

AE\_TD1 = abs(actuals - FTD1)

AE\_TD2 = abs(actuals - FTD2)

AE\_TD3 = abs(actuals - FTD3)

AE\_OP = abs(actuals - FOP)

AE\_GF = abs(actuals - gfcs)

#calculate the Absolute Error for no reconciliation approach

AE\_IF = abs(actuals - fcs)

print(AE\_BU)

print(AE\_TD1)

print(AE\_TD2)

print(AE\_TD3)

print(AE\_OP)

print(AE\_GF)

print(AE\_IF)

print("end of AE")

#calculate the Absolute Percentage Error for each reconciliation approach

APE\_BU = 100\*abs(actuals - FBU) / actuals

APE\_TD1 = 100\*abs(actuals - FTD1) / actuals

APE\_TD2 = 100\*abs(actuals - FTD2) / actuals

APE\_TD3 = 100\*abs(actuals - FTD3) / actuals

APE\_OP = 100\*abs(actuals - FOP) / actuals

APE\_GF = 100\*abs(actuals - gfcs) / actuals

#calculate the Absolute Percentage Error for no reconciliation approach

APE\_IF = 100\*abs(actuals - fcs) / actuals

print(APE\_BU)

print(APE\_TD1)

print(APE\_TD2)

print(APE\_TD3)

print(APE\_OP)

print(APE\_GF)

print(APE\_IF)

print("end of group")

}

for (i in 1:5){

paste("Group",i)

actuals=Actual[Quarter=="Q2" & Group==i]

fcs=IForecast[Quarter=="Q2" & Group==i]

gfcs=GForecast[Quarter=="Q2" & Group==i]

#the reconciled forecasts for each approach

FBU = S %\*% PBU %\*% fcs

FTD1= c(fcs[1], PTD1\*fcs[1])

FTD2= c(fcs[1], PTD2\*fcs[1])

PTD[,1] = fcs[(2):(1+bln)]/sum(fcs[(2):(1+bln)])

FTD3 = S %\*% PTD %\*% fcs

FOP = S %\*% POP %\*% fcs

print(fcs)

print(FBU)

print(FTD1)

print(FTD2)

print(FTD3)

print(FOP)

print(gfcs)

#calculate the Absolute Error for each reconciliation approach

AE\_BU = abs(actuals - FBU)

AE\_TD1 = abs(actuals - FTD1)

AE\_TD2 = abs(actuals - FTD2)

AE\_TD3 = abs(actuals - FTD3)

AE\_OP = abs(actuals - FOP)

AE\_GF = abs(actuals - gfcs)

#calculate the Absolute Error for no reconciliation approach

AE\_IF = abs(actuals - fcs)

print(AE\_BU)

print(AE\_TD1)

print(AE\_TD2)

print(AE\_TD3)

print(AE\_OP)

print(AE\_GF)

print(AE\_IF)

print("end of AE")

#calculate the Absolute Percentage Error for each reconciliation approach

APE\_BU = 100\*abs(actuals - FBU) / actuals

APE\_TD1 = 100\*abs(actuals - FTD1) / actuals

APE\_TD2 = 100\*abs(actuals - FTD2) / actuals

APE\_TD3 = 100\*abs(actuals - FTD3) / actuals

APE\_OP = 100\*abs(actuals - FOP) / actuals

APE\_GF = 100\*abs(actuals - gfcs) / actuals

#calculate the Absolute Percentage Error for no reconciliation approach

APE\_IF = 100\*abs(actuals - fcs) / actuals

print(APE\_BU)

print(APE\_TD1)

print(APE\_TD2)

print(APE\_TD3)

print(APE\_OP)

print(APE\_GF)

print(APE\_IF)

print("end of group")

}

Code for the figures 4.2 (a), (b), (c) and table 4.1

recon=read.table(pipe("pbpaste"), header=T)

attach(recon)

head(recon)

APE\_NR=print(mean(APE[Method=="No\_reconciliation"]))

APE\_BU=print(mean(APE[Method=="Bottom-up"]))

APE\_TD1=print(mean(APE[Method=="Top-Down1"]))

APE\_TD2=print(mean(APE[Method=="Top-Down2"]))

APE\_TD3=print(mean(APE[Method=="Top-Down3"]))

APE\_OP=print(mean(APE[Method=="Optimal\_reconciliation"]))

APE\_GP=print(mean(APE[Method=="Group\_reconciliation"]))

APE\_NR\_Q1=print(mean(APE[Method=="No\_reconciliation" & Quarter=="Q1"]))

APE\_BU\_Q1=print(mean(APE[Method=="Bottom-up" & Quarter=="Q1"]))

APE\_TD1\_Q1=print(mean(APE[Method=="Top-Down1" & Quarter=="Q1"]))

APE\_TD2\_Q1=print(mean(APE[Method=="Top-Down2" & Quarter=="Q1"]))

APE\_TD3\_Q1=print(mean(APE[Method=="Top-Down3" & Quarter=="Q1"]))

APE\_OP\_Q1=print(mean(APE[Method=="Optimal\_reconciliation" & Quarter=="Q1"]))

APE\_GP\_Q1=print(mean(APE[Method=="Group\_reconciliation" & Quarter=="Q1"]))

APE\_NR\_Q2=print(mean(APE[Method=="No\_reconciliation" & Quarter=="Q2"]))

APE\_BU\_Q2=print(mean(APE[Method=="Bottom-up" & Quarter=="Q2"]))

APE\_TD1\_Q2=print(mean(APE[Method=="Top-Down1" & Quarter=="Q2"]))

APE\_TD2\_Q2=print(mean(APE[Method=="Top-Down2" & Quarter=="Q2"]))

APE\_TD3\_Q2=print(mean(APE[Method=="Top-Down3" & Quarter=="Q2"]))

APE\_OP\_Q2=print(mean(APE[Method=="Optimal\_reconciliation" & Quarter=="Q2"]))

APE\_GP\_Q2=print(mean(APE[Method=="Group\_reconciliation" & Quarter=="Q2"]))

APE\_NR\_Apple=print(mean(APE[Method=="No\_reconciliation" & Product=="Apple"]))

APE\_NR\_iPhone=print(mean(APE[Method=="No\_reconciliation" & Product=="iPhone"]))

APE\_NR\_iPad=print(mean(APE[Method=="No\_reconciliation" & Product=="iPad"]))

APE\_NR\_Mac=print(mean(APE[Method=="No\_reconciliation" & Product=="Mac"]))

APE\_BU\_Apple=print(mean(APE[Method=="Bottom-up" & Product=="Apple"]))

APE\_BU\_iPhone=print(mean(APE[Method=="Bottom-up" & Product=="iPhone"]))

APE\_BU\_iPad=print(mean(APE[Method=="Bottom-up" & Product=="iPad"]))

APE\_BU\_Mac=print(mean(APE[Method=="Bottom-up" & Product=="Mac"]))

APE\_TD1\_Apple=print(mean(APE[Method=="Top-Down1" & Product=="Apple"]))

APE\_TD1\_iPhone=print(mean(APE[Method=="Top-Down1" & Product=="iPhone"]))

APE\_TD1\_iPad=print(mean(APE[Method=="Top-Down1" & Product=="iPad"]))

APE\_TD1\_Mac=print(mean(APE[Method=="Top-Down1" & Product=="Mac"]))

APE\_TD2\_Apple=print(mean(APE[Method=="Top-Down2" & Product=="Apple"]))

APE\_TD2\_iPhone=print(mean(APE[Method=="Top-Down2" & Product=="iPhone"]))

APE\_TD2\_iPad=print(mean(APE[Method=="Top-Down2" & Product=="iPad"]))

APE\_TD2\_Mac=print(mean(APE[Method=="Top-Down2" & Product=="Mac"]))

APE\_TD3\_Apple=print(mean(APE[Method=="Top-Down3" & Product=="Apple"]))

APE\_TD3\_iPhone=print(mean(APE[Method=="Top-Down3" & Product=="iPhone"]))

APE\_TD3\_iPad=print(mean(APE[Method=="Top-Down3" & Product=="iPad"]))

APE\_TD3\_Mac=print(mean(APE[Method=="Top-Down3" & Product=="Mac"]))

APE\_OP\_Apple=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Apple"]))

APE\_OP\_iPhone=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPhone"]))

APE\_OP\_iPad=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPad"]))

APE\_OP\_Mac=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Mac"]))

APE\_GP\_Apple=print(mean(APE[Method=="Group\_reconciliation" & Product=="Apple"]))

APE\_GP\_iPhone=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPhone"]))

APE\_GP\_iPad=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPad"]))

APE\_GP\_Mac=print(mean(APE[Method=="Group\_reconciliation" & Product=="Mac"]))

APE\_NR\_Apple\_Q1=print(mean(APE[Method=="No\_reconciliation" & Product=="Apple" & Quarter=="Q1"]))

APE\_NR\_iPhone\_Q1=print(mean(APE[Method=="No\_reconciliation" & Product=="iPhone" & Quarter=="Q1"]))

APE\_NR\_iPad\_Q1=print(mean(APE[Method=="No\_reconciliation" & Product=="iPad" & Quarter=="Q1"]))

APE\_NR\_Mac\_Q1=print(mean(APE[Method=="No\_reconciliation" & Product=="Mac" & Quarter=="Q1"]))

APE\_BU\_Apple\_Q1=print(mean(APE[Method=="Bottom-up" & Product=="Apple" & Quarter=="Q1"]))

APE\_BU\_iPhone\_Q1=print(mean(APE[Method=="Bottom-up" & Product=="iPhone" & Quarter=="Q1"]))

APE\_BU\_iPad\_Q1=print(mean(APE[Method=="Bottom-up" & Product=="iPad" & Quarter=="Q1"]))

APE\_BU\_Mac\_Q1=print(mean(APE[Method=="Bottom-up" & Product=="Mac" & Quarter=="Q1"]))

APE\_TD1\_Apple\_Q1=print(mean(APE[Method=="Top-Down1" & Product=="Apple" & Quarter=="Q1"]))

APE\_TD1\_iPhone\_Q1=print(mean(APE[Method=="Top-Down1" & Product=="iPhone" & Quarter=="Q1"]))

APE\_TD1\_iPad\_Q1=print(mean(APE[Method=="Top-Down1" & Product=="iPad" & Quarter=="Q1"]))

APE\_TD1\_Mac\_Q1=print(mean(APE[Method=="Top-Down1" & Product=="Mac" & Quarter=="Q1"]))

APE\_TD2\_Apple\_Q1=print(mean(APE[Method=="Top-Down2" & Product=="Apple" & Quarter=="Q1"]))

APE\_TD2\_iPhone\_Q1=print(mean(APE[Method=="Top-Down2" & Product=="iPhone" & Quarter=="Q1"]))

APE\_TD2\_iPad\_Q1=print(mean(APE[Method=="Top-Down2" & Product=="iPad" & Quarter=="Q1"]))

APE\_TD2\_Mac\_Q1=print(mean(APE[Method=="Top-Down2" & Product=="Mac" & Quarter=="Q1"]))

APE\_TD3\_Apple\_Q1=print(mean(APE[Method=="Top-Down3" & Product=="Apple" & Quarter=="Q1"]))

APE\_TD3\_iPhone\_Q1=print(mean(APE[Method=="Top-Down3" & Product=="iPhone" & Quarter=="Q1"]))

APE\_TD3\_iPad\_Q1=print(mean(APE[Method=="Top-Down3" & Product=="iPad" & Quarter=="Q1"]))

APE\_TD3\_Mac\_Q1=print(mean(APE[Method=="Top-Down3" & Product=="Mac" & Quarter=="Q1"]))

APE\_OP\_Apple\_Q1=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Apple" & Quarter=="Q1"]))

APE\_OP\_iPhone\_Q1=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPhone" & Quarter=="Q1"]))

APE\_OP\_iPad\_Q1=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPad" & Quarter=="Q1"]))

APE\_OP\_Mac\_Q1=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Mac" & Quarter=="Q1"]))

APE\_GP\_Apple\_Q1=print(mean(APE[Method=="Group\_reconciliation" & Product=="Apple" & Quarter=="Q1"]))

APE\_GP\_iPhone\_Q1=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPhone" & Quarter=="Q1"]))

APE\_GP\_iPad\_Q1=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPad" & Quarter=="Q1"]))

APE\_GP\_Mac\_Q1=print(mean(APE[Method=="Group\_reconciliation" & Product=="Mac" & Quarter=="Q1"]))

APE\_NR\_Apple\_Q2=print(mean(APE[Method=="No\_reconciliation" & Product=="Apple" & Quarter=="Q2"]))

APE\_NR\_iPhone\_Q2=print(mean(APE[Method=="No\_reconciliation" & Product=="iPhone" & Quarter=="Q2"]))

APE\_NR\_iPad\_Q2=print(mean(APE[Method=="No\_reconciliation" & Product=="iPad" & Quarter=="Q2"]))

APE\_NR\_Mac\_Q2=print(mean(APE[Method=="No\_reconciliation" & Product=="Mac" & Quarter=="Q2"]))

APE\_BU\_Apple\_Q2=print(mean(APE[Method=="Bottom-up" & Product=="Apple" & Quarter=="Q2"]))

APE\_BU\_iPhone\_Q2=print(mean(APE[Method=="Bottom-up" & Product=="iPhone" & Quarter=="Q2"]))

APE\_BU\_iPad\_Q2=print(mean(APE[Method=="Bottom-up" & Product=="iPad" & Quarter=="Q2"]))

APE\_BU\_Mac\_Q2=print(mean(APE[Method=="Bottom-up" & Product=="Mac" & Quarter=="Q2"]))

APE\_TD1\_Apple\_Q2=print(mean(APE[Method=="Top-Down1" & Product=="Apple" & Quarter=="Q2"]))

APE\_TD1\_iPhone\_Q2=print(mean(APE[Method=="Top-Down1" & Product=="iPhone" & Quarter=="Q2"]))

APE\_TD1\_iPad\_Q2=print(mean(APE[Method=="Top-Down1" & Product=="iPad" & Quarter=="Q2"]))

APE\_TD1\_Mac\_Q2=print(mean(APE[Method=="Top-Down1" & Product=="Mac" & Quarter=="Q2"]))

APE\_TD2\_Apple\_Q2=print(mean(APE[Method=="Top-Down2" & Product=="Apple" & Quarter=="Q2"]))

APE\_TD2\_iPhone\_Q2=print(mean(APE[Method=="Top-Down2" & Product=="iPhone" & Quarter=="Q2"]))

APE\_TD2\_iPad\_Q2=print(mean(APE[Method=="Top-Down2" & Product=="iPad" & Quarter=="Q2"]))

APE\_TD2\_Mac\_Q2=print(mean(APE[Method=="Top-Down2" & Product=="Mac" & Quarter=="Q2"]))

APE\_TD3\_Apple\_Q2=print(mean(APE[Method=="Top-Down3" & Product=="Apple" & Quarter=="Q2"]))

APE\_TD3\_iPhone\_Q2=print(mean(APE[Method=="Top-Down3" & Product=="iPhone" & Quarter=="Q2"]))

APE\_TD3\_iPad\_Q2=print(mean(APE[Method=="Top-Down3" & Product=="iPad" & Quarter=="Q2"]))

APE\_TD3\_Mac\_Q2=print(mean(APE[Method=="Top-Down3" & Product=="Mac" & Quarter=="Q2"]))

APE\_OP\_Apple\_Q2=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Apple" & Quarter=="Q2"]))

APE\_OP\_iPhone\_Q2=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPhone" & Quarter=="Q2"]))

APE\_OP\_iPad\_Q2=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="iPad" & Quarter=="Q2"]))

APE\_OP\_Mac\_Q2=print(mean(APE[Method=="Optimal\_reconciliation" & Product=="Mac" & Quarter=="Q2"]))

APE\_GP\_Apple\_Q2=print(mean(APE[Method=="Group\_reconciliation" & Product=="Apple" & Quarter=="Q2"]))

APE\_GP\_iPhone\_Q2=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPhone" & Quarter=="Q2"]))

APE\_GP\_iPad\_Q2=print(mean(APE[Method=="Group\_reconciliation" & Product=="iPad" & Quarter=="Q2"]))

APE\_GP\_Mac\_Q2=print(mean(APE[Method=="Group\_reconciliation" & Product=="Mac" & Quarter=="Q2"]))

**Appendix C**

**R code for calculation of AvgRelMAE**

data=read.table(pipe("pbpaste"), header=T)

attach(data)

head(data)

#ni for each time series is 10 (5 groups and 2 quarters)

n=10

#total ni for each method is 40 (corresponding to 4 time-series)

N=40

MAEGR\_Apple=mean(AE[Product=="Apple" & Method=="Group\_reconciliation"])

MAEGR\_iPhone=mean(AE[Product=="iPhone" & Method=="Group\_reconciliation"])

MAEGR\_iPad=mean(AE[Product=="iPad" & Method=="Group\_reconciliation"])

MAEGR\_Mac=mean(AE[Product=="Mac" & Method=="Group\_reconciliation"])

#comparison with Bottom-up method

MAEBU\_Apple=mean(AE[Product=="Apple" & Method=="Bottom-up"])

MAEBU\_iPhone=mean(AE[Product=="iPhone" & Method=="Bottom-up"])

MAEBU\_iPad=mean(AE[Product=="iPad" & Method=="Bottom-up"])

MAEBU\_Mac=mean(AE[Product=="Mac" & Method=="Bottom-up"])

r\_GR\_BU\_Apple=MAEGR\_Apple/MAEBU\_Apple

r\_GR\_BU\_iPhone=MAEGR\_iPhone/MAEBU\_iPhone

r\_GR\_BU\_iPad=MAEGR\_iPad/MAEBU\_iPad

r\_GR\_BU\_Mac=MAEGR\_Mac/MAEBU\_Mac

AvgRelMAE\_GR\_BU=((r\_GR\_BU\_Apple^n)\*(r\_GR\_BU\_iPhone^n)\*(r\_GR\_BU\_iPad^n)\*(r\_GR\_BU\_Mac^n))^(1/N)

AvgRelMAE\_GR\_BU

#comparison with Top-Down (Average historical proportions) method

MAETD1\_Apple=mean(AE[Product=="Apple" & Method=="Top-Down1"])

MAETD1\_iPhone=mean(AE[Product=="iPhone" & Method=="Top-Down1"])

MAETD1\_iPad=mean(AE[Product=="iPad" & Method=="Top-Down1"])

MAETD1\_Mac=mean(AE[Product=="Mac" & Method=="Top-Down1"])

r\_GR\_TD1\_Apple=MAEGR\_Apple/MAETD1\_Apple

r\_GR\_TD1\_iPhone=MAEGR\_iPhone/MAETD1\_iPhone

r\_GR\_TD1\_iPad=MAEGR\_iPad/MAETD1\_iPad

r\_GR\_TD1\_Mac=MAEGR\_Mac/MAETD1\_Mac

AvgRelMAE\_GR\_TD1=((r\_GR\_TD1\_Apple^n)\*(r\_GR\_TD1\_iPhone^n)\*(r\_GR\_TD1\_iPad^n)\*(r\_GR\_TD1\_Mac^n))^(1/N)

AvgRelMAE\_GR\_TD1

#comparison with Top-Down (Proportions of historical averages) method

MAETD2\_Apple=mean(AE[Product=="Apple" & Method=="Top-Down2"])

MAETD2\_iPhone=mean(AE[Product=="iPhone" & Method=="Top-Down2"])

MAETD2\_iPad=mean(AE[Product=="iPad" & Method=="Top-Down2"])

MAETD2\_Mac=mean(AE[Product=="Mac" & Method=="Top-Down2"])

r\_GR\_TD2\_Apple=MAEGR\_Apple/MAETD2\_Apple

r\_GR\_TD2\_iPhone=MAEGR\_iPhone/MAETD2\_iPhone

r\_GR\_TD2\_iPad=MAEGR\_iPad/MAETD2\_iPad

r\_GR\_TD2\_Mac=MAEGR\_Mac/MAETD2\_Mac

AvgRelMAE\_GR\_TD2=((r\_GR\_TD2\_Apple^n)\*(r\_GR\_TD2\_iPhone^n)\*(r\_GR\_TD2\_iPad^n)\*(r\_GR\_TD2\_Mac^n))^(1/N)

AvgRelMAE\_GR\_TD2

#comparison with Top-Down (Forecasts proportions) method

MAETD3\_Apple=mean(AE[Product=="Apple" & Method=="Top-Down3"])

MAETD3\_iPhone=mean(AE[Product=="iPhone" & Method=="Top-Down3"])

MAETD3\_iPad=mean(AE[Product=="iPad" & Method=="Top-Down3"])

MAETD3\_Mac=mean(AE[Product=="Mac" & Method=="Top-Down3"])

r\_GR\_TD3\_Apple=MAEGR\_Apple/MAETD3\_Apple

r\_GR\_TD3\_iPhone=MAEGR\_iPhone/MAETD3\_iPhone

r\_GR\_TD3\_iPad=MAEGR\_iPad/MAETD3\_iPad

r\_GR\_TD3\_Mac=MAEGR\_Mac/MAETD3\_Mac

AvgRelMAE\_GR\_TD3=((r\_GR\_TD3\_Apple^n)\*(r\_GR\_TD3\_iPhone^n)\*(r\_GR\_TD3\_iPad^n)\*(r\_GR\_TD3\_Mac^n))^(1/N)

AvgRelMAE\_GR\_TD3

#comparison with Optimal\_reconciliation method

MAEOR\_Apple=mean(AE[Product=="Apple" & Method=="Optimal\_reconciliation"])

MAEOR\_iPhone=mean(AE[Product=="iPhone" & Method=="Optimal\_reconciliation"])

MAEOR\_iPad=mean(AE[Product=="iPad" & Method=="Optimal\_reconciliation"])

MAEOR\_Mac=mean(AE[Product=="Mac" & Method=="Optimal\_reconciliation"])

r\_GR\_OR\_Apple=MAEGR\_Apple/MAEOR\_Apple

r\_GR\_OR\_iPhone=MAEGR\_iPhone/MAEOR\_iPhone

r\_GR\_OR\_iPad=MAEGR\_iPad/MAEOR\_iPad

r\_GR\_OR\_Mac=MAEGR\_Mac/MAEOR\_Mac

AvgRelMAE\_GR\_OR=((r\_GR\_OR\_Apple^n)\*(r\_GR\_OR\_iPhone^n)\*(r\_GR\_OR\_iPad^n)\*(r\_GR\_OR\_Mac^n))^(1/N)

AvgRelMAE\_GR\_OR

**Appendix D**

**R code for deviation evaluation**

data=read.table(pipe("pbpaste"),header=T)

head(data)

attach(data)

mean(PDeviation)

mean(PDeviation[Product=="Apple"])

mean(PDeviation[Product!="Apple"])

boxplot(PDeviation~Product)

boxplot(Deviation~Product)

boxplot(PDeviation~Product)

max(PDeviation)

min(PDeviation)

table(PDeviation == "0" & Product=="Apple")

table(PDeviation == "0" & Product!="Apple")

table(PDeviation < "0")

table(PDeviation > "0")

table(Group[PDeviation==0])

par(mfrow=c(2,1))

boxplot(PDeviation~Product)

boxplot(PDeviation~Group)

wilcox.test(IForecast,GForecast,paired=TRUE)

**Appendix E**

**R code for comparison between deviation and forecast performance**

data=read.table(pipe("pbpaste"),header=T)

attach(data)

cor.test(Deviation, D\_APE,method="spearman")

DAPE\_apple=D\_APE[Product=="Apple"]

D\_apple=Deviation[Product=="Apple"]

cor.test(D\_apple, DAPE\_apple, method="spearman")

D\_o=Deviation[Product!="Apple"]

DAPE\_o=D\_APE[Product!="Apple"]

cor.test(D\_o, DAPE\_o, method="spearman")

D\_C=Deviation[Group=="C"]

DAPE\_C=D\_APE[Group=="C"]

cor.test(D\_C,DAPE\_C, method="spearman")

D\_D=Deviation[Group=="D"]

DAPE\_D=D\_APE[Group=="D"]

cor.test(D\_D,DAPE\_D, method="spearman")

D\_E=Deviation[Group=="E"]

DAPE\_E=D\_APE[Group=="E"]

cor.test(D\_E,DAPE\_E, method="spearman")

DAPE\_F=D\_APE[Group=="F"]

D\_F=Deviation[Group=="F"]

cor.test(D\_F,DAPE\_F, method="spearman")

D\_G=Deviation[Group=="G"]

DAPE\_G=D\_APE[Group=="G"]

cor.test(D\_G,DAPE\_G, method="spearman")