



**HAL**  
open science

## ARIMA demand forecasting by aggregation

Bahman Rostami Tabar

► **To cite this version:**

Bahman Rostami Tabar. ARIMA demand forecasting by aggregation. Other. Université Sciences et Technologies - Bordeaux I, 2013. English. NNT : 2013BOR15228 . tel-00980614

**HAL Id: tel-00980614**

**<https://theses.hal.science/tel-00980614>**

Submitted on 18 Apr 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE**  
**PRÉSENTÉE À**  
**L'UNIVERSITÉ BORDEAUX 1**

ÉCOLE DOCTORALE DES SCIENCES PHYSIQUES ET DE L'INGENIEUR

Par Bahman ROSTAMI-TABAR  
POUR OBTENIR LE GRADE DE  
DOCTEUR  
SPÉCIALITÉ: PRODUCTIQUE

**ARIMA DEMAND FORECASTING BY AGGREGATION**

Soutenue le: 10 Décembre 2013

Devant la commission d'examen formée de:

Alexandre DOLGUI	Professeur, Ecole des Mines de Saint-Etienne	Rapporteur
Yannick FREIN	Professeur, INP Grenoble	Rapporteur
Konstantinos NIKOLOPOULOS	Professeur, Bangor University, Royaume-Uni	Examineur
Jean-Paul BOURRIERES	Professeur, Université Bordeaux 1	Examineur
Aris SYNTETOS	Professeur, Cardiff University, Royaume-Uni	Invité
Mohamed Zied BABAI	Enseignant-Chercheur, Kedge Business school	Co-Directeur de thèse
Yves DUCQ	Professeur, Université Bordeaux 1	Directeur de thèse

## Abstract

Demand forecasting performance is subject to the uncertainty underlying the time series an organisation is dealing with. There are many approaches that may be used to reduce demand uncertainty and consequently improve the forecasting (and inventory control) performance. An intuitively appealing such approach that is known to be effective is demand aggregation. One approach is to aggregate demand in lower-frequency ‘time buckets’. Such an approach is often referred to, in the academic literature, as temporal aggregation. Another approach discussed in the literature is that associated with cross-sectional aggregation, which involves aggregating different time series to obtain higher level forecasts.

This research discusses whether it is appropriate to use the original (not aggregated) data to generate a forecast or one should rather aggregate data first and then generate a forecast. This Ph.D. thesis reveals the conditions under which each approach leads to a superior performance as judged based on forecast accuracy. Throughout this work, it is assumed that the underlying structure of the demand time series follows an AutoRegressive Integrated Moving Average (ARIMA) process.

In the first part of our<sup>1</sup> research, the effect of temporal aggregation on demand forecasting is analysed. It is assumed that the non-aggregate demand follows an autoregressive moving average process of order one, ARMA(1,1). Additionally, the associated special cases of a first-order autoregressive process, AR(1) and a moving average process of order one, MA(1) are also considered, and a Single Exponential Smoothing (SES) procedure is used to forecast demand. These demand processes are often encountered in practice and SES is one of the standard estimators used in industry. Theoretical Mean Squared Error expressions are derived for the aggregate and the non-aggregate demand in order to contrast the relevant forecasting performances. The theoretical analysis is validated by an extensive numerical investigation and experimentation with an empirical dataset. The results indicate that performance improvements achieved through the aggregation approach are a

---

<sup>1</sup> The use of the words “our” and “we” throughout the thesis is purely conventional. The work presented in this Ph.D. thesis is the result of research conducted by the author alone, albeit with support from an academic institution and a supervisory team.

function of the aggregation level, the smoothing constant value used for SES and the process parameters.

In the second part of our research, the effect of cross-sectional aggregation on demand forecasting is evaluated. More specifically, the relative effectiveness of top-down (TD) and bottom-up (BU) approaches are compared for forecasting the aggregate and sub-aggregate demands. It is assumed that the sub-aggregate demand follows either a ARMA(1,1) or a non-stationary Integrated Moving Average process of order one, IMA(1,1) and a SES procedure is used to extrapolate future requirements. Such demand processes are often encountered in practice and, as discussed above, SES is one of the standard estimators used in industry (in addition to being the optimal estimator for an IMA(1) process). Theoretical Mean Squared Errors are derived for the BU and TD approach in order to contrast the relevant forecasting performances. The theoretical analysis is supported by an extensive numerical investigation at both the aggregate and sub-aggregate levels in addition to empirically validating our findings on a real dataset from a European superstore. The results show that the superiority of each approach is a function of the series autocorrelation, the cross-correlation between series and the comparison level.

Finally, for both parts of the research, valuable insights are offered to practitioners and an agenda for further research in this area is provided.

Keywords: demand forecasting; temporal aggregation; cross-sectional aggregation; stationary processes; nonstationary processes; single exponential smoothing

## **Acknowledgements**

I am grateful to most merciful and ALMIGHTY “ALLAH” who gave me the health and opportunity to complete this thesis.

I am delighted to place on record my deepest sense of gratitude and profound obligation to my supervisor Prof Yves Ducq for his endless support, guidance and encouragement during my three years as a PhD student. I appreciate all his contributions of time and ideas.

I am deeply indebted to my supervisor, Dr. Mohamed-Zied Babai for his advice and patience that stimulated and nourished my intellectual maturity. His never ending energy and enthusiasm for research made this PhD inspiring, challenging and very interesting.

I would like to express my deep and sincere gratitude to Prof. Aris Syntetos. His in-depth knowledge and particularly his advice have been invaluable in ensuring I was on the right track throughout the Ph.D. studies.

I am grateful to the Campus France for the scholarship they provided for this Ph.D study. I am also obliged to BEM (Bordeaux Management School) for their support that enabled me to attend and present my academic papers at international conferences.

My sincere thanks to my brother, sisters, brothers-in-law, nephews and nieces for their love and encouragement and for believing in me.

Words fail to express my appreciation for my beloved wife Rojin whose eternal love, understating, care but most of all confidence in me took this doctoral study towards completion.

Last but not least, my deepest gratitude to my parents Ibrahim and Afrooz, whose lessons on hardworking and diligence enabled me to complete my PhD thesis.

**To my mother who passed away in August 2008**

**To my Father**

**To Rojin: My best Friend**

# Contents

<b>List of Appendices</b> -----	vii
<b>List of Figures</b> -----	viii
<b>List of Tables</b> -----	x
<b>Chapter 1 Introduction and Problem Statement</b> -----	<b>13</b>
1.1 Definitions-----	13
1.2 Business Context-----	22
1.3 Research Background-----	25
1.4 Research Overview-----	27
1.5 Methodology-----	29
1.6 Thesis Structure-----	29
<b>Chapter 2 State of the art</b> -----	<b>31</b>
2.1 Introduction-----	31
2.2 Temporal Aggregation-----	33
2.2.1 Temporal aggregation identification process-----	33
2.2.2 Demand forecasting by temporal aggregation-----	40
2.3 Cross-sectional aggregation-----	43
2.3.1 Cross-sectional aggregation identification process-----	44
2.3.2 Demand forecasting by cross-sectional aggregation-----	45
2.4 Discussion on the literature review-----	47
<b>Chapter 3 Temporal aggregation</b> -----	<b>50</b>
3.1 Theoretical Analysis-----	52
3.1.1 Notation and assumptions-----	53

3.1.2	MSE derivation at disaggregate level	64
3.1.2.1	MSE Before Aggregation, $MSE_{BA}$	64
3.1.2.2	MSE after Aggregation, $MSE_{AA}$	67
3.1.3	MSE derivation at aggregate level	70
3.1.3.1	MSE Before Aggregation, $MSE_{BA}$	70
3.1.3.2	MSE after Aggregation, $MSE_{AA}$	72
3.2	Comparative analysis	74
3.2.1	Impact of the parameters – sensitivity analysis	74
3.2.1.1	Comparison at disaggregate level	75
3.2.1.2	Comparison at aggregate level	86
3.2.2	Theoretical Comparison	95
3.2.2.1	Comparison at disaggregate level	95
3.2.2.2	Comparison at aggregate level	101
3.2.3	Optimal aggregation level	107
3.2.3.1	Comparison at disaggregate level	107
3.2.3.2	Comparison at aggregate level	110
3.3	Simulation investigation	112
3.3.1	Simulation design	113
3.3.2	Simulation Result	114
3.4	Empirical analysis	118
3.4.1	Empirical Dataset and Experiment Details	118
3.4.2	Empirical Results	121
3.4.2.1	Comparison at disaggregate level	121
3.4.2.2	Comparison at aggregate level	125
3.5	Conclusion	129
<b>Chapter 4</b>	<b>Cross-Sectional Aggregation</b>	<b>130</b>
4.1	Theoretical analysis	132
4.1.1	Notation and assumptions	133
4.1.2	Variance of forecast error at aggregate level	139
4.1.2.1	Integrated moving average process order one, ARIMA(0,1,1)	139

4.1.2.2	Autoregressive moving average process order one, ARIMA(1,0,1)-----	141
4.1.3	Variance of forecast error at subaggregate level -----	144
4.1.3.1	Autoregressive moving average process order one, ARIMA(1,0,1)-----	145
4.2	Theoretical Comparison -----	147
4.2.1	Comparison at aggregate level-----	147
4.2.1.1	Integrated moving average process order one ARIMA(0,1,1)-----	147
4.2.1.2	Autoregressive moving average process order one ARIMA(1,0,1)-----	148
4.2.2	Comparison at subaggregate level-----	149
4.2.2.1	Autoregressive moving average process order one ARIMA(1,0,1)-----	149
4.3	Simulation study -----	151
4.3.1	Simulation design -----	151
4.3.2	Simulation results -----	153
4.3.2.1	Comparison at the Aggregate Level -----	153
4.3.2.2	Comparison at Subaggregate Level-----	159
4.4	Empirical analysis-----	163
4.4.1	Empirical dataset and experiment details-----	163
4.4.2	Empirical results-----	165
4.5	Conclusion-----	167
<b>Chapter 5 Conclusions and Future Research-----</b>		<b>170</b>
5.1	Contributions of the Thesis-----	171
5.1.1	Temporal aggregation -----	171
5.1.2	Cross sectional aggregation -----	173
5.2	Managerial implications-----	174
5.2.1	Temporal aggregation -----	174
5.2.2	Cross sectional aggregation -----	175
5.3	Limitations and future research -----	176
<b>References -----</b>		<b>179</b>

## List of Appendices

Appendix A: The relationship of autocovariance between nonaggregate and aggregate demand .....	186
Appendix B: Covariance between the disaggregate demand and aggregate forecast for ARIMA(1,0,1) process .....	194
Appendix C: Variance of the aggregate forecast for the ARIMA(1,0,1) process .....	195
Appendix D: Covariance of the aggregate demand and non-aggregate forecast for ARIMA(1,0,1) process .....	196
Appendix E: Coefficient of variation before and after aggregation for ARIMA(1,0,1) .....	197
Appendix F: Proof of theorem 1-3 .....	198
Appendix G: Selection procedure for the ARIMA(1,0,1) process.....	201
Appendix H: Proof of theorem 2-3, ARIMA(0,0,1) .....	202
Appendix I: Selection procedure for the ARIMA(1,0,0) process .....	204
Appendix J: Proof of theorem 3-3 and theorem 4-3.....	205
Appendix K: Selection procedure for the ARIMA(1,0,1) process- Comparison at the aggregate level.....	206
Appendix L: Proof of theorem 5-3 .....	208
Appendix M: Proof of theorem 7-3 for ARIMA(0,0,1) – Comparison at the aggregate level	209
Appendix N: Proof of theorem 9-3 for an ARIMA(1,0,0) - Comparison at aggregate level..	210
Appendix O: Proof of theorem 9-3 for an ARIMA(1,0,0) - Comparison at aggregate level..	210
Appendix P: Covariance between demand $i,j$ and forecast $i,j$ .....	213
Appendix Q: Proof of theorem 4-1 .....	214
Appendix R: Proof of theorem 4-2, collary 4-1 and collary 4-2 .....	215

## List of Figures

<i>Figure 1-1: Non-overlapping temporal aggregation (from weekly to monthly data)</i> .....	18
<i>Figure 1-2: Overlapping temporal aggregation (from weekly to monthly data)</i> .....	19
<i>Figure 1-3: Comparison at disaggregate level</i> .....	20
<i>Figure 1-4: Comparison at aggregate level</i> .....	20
<i>Figure 1-5: Schematic diagram of TD (left) and BU (right) approaches</i> .....	21
<i>Figure 1-6: Research Overview</i> .....	28
<i>Figure 1-7: Methodology</i> .....	29
<i>Figure 3-1: Sample autocorrelation of ARIMA(1,0,1) process when <math>\theta = -0.8</math> and <math>\phi = 0.9</math></i> .....	56
<i>Figure 3-2: Sample autocorrelation of ARIMA(1,0,1) process when <math>\theta = 0.8</math> and <math>\phi = -0.7</math></i> .....	57
<i>Figure 3-3: Autocorrelation of ARIMA(0,0,1) process</i> .....	58
<i>Figure 3-4: Sample autocorrelation and the process shape of ARIMA(0,0,1) process when <math>\theta = -0.9</math></i> .....	59
<i>Figure 3-5: Sample autocorrelation and the process shape of ARIMA(0,0,1) process when <math>\theta = 0.9</math></i> .....	60
<i>Figure 3-6: Sample autocorrelation and the process shape of ARIMA(1,0,0) process when <math>\phi = 0.9</math></i> .....	62
<i>Figure 3-7: Sample autocorrelation and the process shape of ARIMA(1,0,0) process when <math>\phi = -0.9</math></i> .....	63
<i>Figure 3-8: Impact of <math>m</math>, <math>\theta</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the ratio of MSE: <math>\alpha = 0.1</math>, <math>\beta = 0.01</math> (top)</i> .....	76
<i>Figure 3-9: Impact of <math>m</math>, <math>\theta</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the ratio of MSE: <math>\alpha = 0.3</math>, <math>\beta = 0.1</math> (top)</i> .....	77
<i>Figure 3-10: Impact of <math>m</math>, <math>\theta</math>, <math>\alpha</math> and <math>\beta</math> on the MSE ratio for <math>\alpha = 0.1</math> (top) and <math>\alpha = 0.5</math> (bottom)</i> .....	80
<i>Figure 3-11: Impact of control parameters for ARIMA(0,0,1) process on the MSE ratio</i> .....	81
<i>Figure 3-12: Impact of <math>m</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the MSE ratio for <math>\alpha = 0.1</math> (top) and <math>\alpha = 0.5</math> (bottom)</i> .....	84
<i>Figure 3-13: Impact of control parameters for ARIMA(1,0,0) process on the MSE ratio</i> .....	85
<i>Figure 3-14: Impact of <math>m</math>, <math>\theta</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the ratio of MSE: <math>\alpha = 0.1</math>, <math>\beta = 0.01</math> (top)</i> <i><math>\alpha = 0.1</math>, <math>\beta = 0.05</math> (bottom)</i> .....	87
<i>Figure 3-15: Impact of <math>m</math>, <math>\theta</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the ratio of MSE: <math>\alpha = 0.3</math>, <math>\beta = 0.1</math> (top)</i> .....	88
<i>Figure 3-16: Impact of <math>m</math>, <math>\theta</math>, <math>\alpha</math> and <math>\beta</math> on the MSE ratio for <math>\alpha = 0.1</math> (top) and <math>\alpha = 0.5</math> (bottom)</i> .....	90
<i>Figure 3-17: Impact of control parameters for ARIMA(0,0,1) process on the MSE ratio</i> .....	91
<i>Figure 3-18: Impact of <math>m</math>, <math>\phi</math>, <math>\alpha</math> and <math>\beta</math> on the MSE ratio for <math>\alpha = 0.1</math> (top) and <math>\alpha = 0.5</math> (bottom)</i> .....	93
<i>Figure 3-19: Impact of control parameters for ARIMA(1,0,0) process on the MSE ratio</i> .....	94
<i>Figure 3-20: MSE ratio for different values of <math>m</math> for an ARIMA(1,0,1) process</i> .....	108
<i>Figure 3-21: MSE ratio for different values of <math>m</math> ARIMA(1,0,0) process at disaggregate level</i> .....	109

<b>Figure 3-22:</b> MSE for different values of $m$ , ARIMA(1,0,1) process compared at aggregate level.....	111
<b>Figure 3-23:</b> MSE for different values of $m$ , ARIMA(0,0,1) process comparison at aggregate level.....	112
<b>Figure 3-24:</b> MSE for different values of $m$ , ARIMA(1,0,0) process comparison at aggregate level.....	112
<b>Figure 3-25:</b> Impact of $m$ , $\theta$ , $\phi$ , $\alpha$ and $\beta$ on the ratio of MSE compared at disaggregate level: $\alpha = 0.1, \beta = 0.01$ (top), $\alpha = 0.1, \beta = 0.05$ (bottom).....	116
<b>Figure 3-26:</b> Impact of $m$ , $\theta$ , $\phi$ , $\alpha$ and $\beta$ on the ratio of MSE compared at aggregate level: $\alpha = 0.1, \beta = 0.01$ (top), $\alpha = 0.1, \beta = 0.05$ (bottom).....	117
<b>Figure 3-27:</b> Empirical results compared at disaggregate level, ARIMA(1,0,1) process .....	122
<b>Figure 3-28:</b> Cut-off points of $\beta$ implying an outperformance of the aggregation approach for different values of $\alpha$ and $m$ compared at disaggregate level, ARIMA(0,0,1) process .....	123
<b>Figure 3-29:</b> Cut-off points of $\beta$ implying an outperformance of the aggregation approach for different values of $\alpha$ and $m$ compared at disaggregate level, ARIMA(1,0,0) process with $-1 < \phi \leq 0.33$ . .....	124
<b>Figure 3-30:</b> Empirical results compared at aggregate level, ARIMA(1,0,1) process .....	126
<b>Figure 3-31:</b> Cut-off points of $\beta$ implying an outperformance of the aggregation approach for different values of $\alpha$ and $m$ compared at aggregate level, ARIMA(0,0,1) process. ....	126
<b>Figure 3-32:</b> Cut-off points of $\beta$ implying an outperformance of the aggregation approach for different values of $\alpha$ and $m$ compared at aggregate level, ARIMA(1,0,0) process with $-1 < \phi < 0.33$ . ....	127
<b>Figure 3-33:</b> Cut-off points of $\beta$ implying an outperformance of the aggregation approach for different values of $\alpha$ and $m$ compared at aggregate level, ARIMA(1,0,0) process with $0.33 < \phi < 1$ . ....	127
<b>Figure 4-1:</b> Sample autocorrelation of ARIMA(0,1,1) process when $\theta = -0.9$ . ....	135
<b>Figure 4-2:</b> Sample autocorrelation of ARIMA(1,0,1) process when $\theta = 0.9$ . ....	136
<b>Figure 4-3:</b> Relative performance of the TD and the BU approaches in forecasting aggregate demand under different combinations of $\theta_1$ , $\theta_2$ and $\rho_{12}$ for an ARIMA(0,1,1) demand process. ....	155
<b>Figure 4-4 :</b> Relative performance of the TD and the BU approaches in forecasting aggregate demand under different combinations of $\theta_1$ , $\theta_2$ and $\rho_{12}$ for an ARIMA(1,0,1) process. ....	156
<b>Figure 4-5:</b> Relative performance of TD and BU approaches in forecasting subaggregate items under different values of $\theta_1$ , $\theta_2$ , $\rho_{12}$ for an ARIMA(0,1,1) process. ....	160
<b>Figure 4-6 :</b> Relative performance of TD and BU approaches in forecasting subaggregate items under different values of $\theta_1$ , $\theta_2$ , $\rho_{12}$ for an ARIMA(1,0,1) process. ....	161

## List of Tables

<i>Table 2-1: the effect of the non-overlapping temporal aggregation on process's structure</i> .....	36
<i>Table 2-2: the effect of the overlapping temporal aggregation on process's structure</i> .....	39
<i>Table 2-3: the effect of temporal aggregation on integer ARIMA type process's structure</i> .....	39
<i>Table 2-4: aggregate process of cross-sectional aggregation</i> .....	44
<i>Table 3-1: Autocorrelation of ARIMA(1,0,1) process</i> .....	55
<i>Table 3-2: Selection procedure for the ARIMA(1,0,1) process, Comparison at disaggregate level</i> .....	97
<i>Table 3-3: Selection procedure for the ARIMA(1,0,1) process, Comparison at aggregate level</i> .....	102
<i>Table 3-4: Parameters of the simulation experiment</i> .....	113
<i>Table 3-5: Processes present in the empirical data set, ARIMA(1,0,1) process</i> .....	119
<i>Table 3-6: Processes Present in the Empirical Data Set, ARIMA(0,0,1) process</i> .....	119
<i>Table 3-7: Processes Present in the Empirical Data Set, ARIMA(1,0,0) process</i> .....	120
<i>Table 4-1: The ratio of <math>V_{TD}/V_{BU}</math> for different control parameters and <math>-1 &lt; \phi_1 = \phi_2 \leq .5</math>, <math>-1 &lt; \theta_1 = \theta_2 &lt; 1</math>..</i>	150
<i>Table 4-2: The ratio of <math>V_{TD}/V_{BU}</math> for different control parameters and <math>0.5 &lt; \phi_1 = \phi_2 &lt; 1</math>, <math>0 &lt; \theta_1 = \theta_2 &lt; 1</math></i>	150
<i>Table 4-3: Parameters of the simulation experiment</i> .....	152
<i>Table 4-4: The empirical data set for ARIMA(0,1,1)</i> .....	164
<i>Table 4-5: The empirical data set for ARIMA(1,0,1)</i> .....	164
<i>Table 4-6: The empirical ratio of <math>V_{TD}/V_{BU}</math> for an ARIMA(0,1,1) process</i> .....	165
<i>Table 4-7: The empirical ratio of <math>V_{TD}/V_{BU}</math> by considering the aggregation between different groups (intervals of <math>\theta</math> values)</i> .....	166
<i>Table 4-8: The empirical results for ARIMA(1,0,1)</i> .....	167

## Chapter 1 Introduction and Problem Statement

---

This chapter provides the overall academic perspective, the objectives of this work and the steps required to conduct the research and meet the objectives. First, some key terms in the area of demand aggregation and forecasting are defined. Then, the business context, the research background and an overview of the research and its objectives are presented before discussing the methodological approach employed for the purposes of this work. We elaborate on all these issues later on in the thesis in chapters 2 - 4. The structure of this PhD thesis is presented at the end of the chapter.

To attain a unified understanding of concepts related to this research work, it is necessary to take a step back and provide the definition of some key terms.

### 1. Definitions

In this section, a brief description of the key terms and phrases used in this research work is provided. These are the terms that are being used all along this thesis and specifically in chapters 3 and 4.

- **Time series**

Makridakis et al (1998) defined a time series as a sequence of observations over time. A time series is an ordered sequence of observations. Although, the order is usually through time, particularly in terms of some equally spaced time intervals, the ordering may also be taken through other dimensions, such as space (Harvey, 1993). Time series occur in a variety of fields such as agriculture, business and economics, engineering, geophysics, medical science, social science, etc. For example in the business context, annual production levels, monthly spare parts demand, weekly inventory levels and daily sales all constitute examples of time series. In this thesis, we focus on (weekly) demand time series. With regards to the

empirical data used for the purposes of our research an important qualification needs to be made. Sales figures are being used as a proxy for demand. Demand itself may not necessarily equal the sales, in case of requests not being satisfied due to stock outs. That is, demand would equal the (achieved) sales plus the lost (or backordered) sales. However, it is reasonable to use this approximation and a necessary condition.

- **Stationary time series**

A stationary time series is one whose properties do not depend on the time at which the series is observed (Makridakis et al., 1998). For a stochastic process to be stationary the expected value of the time series, the variance and the autocovariance of any lag  $k$  does not depend on time (Harvey, 1993). The most general class of stationary models for forecasting a time series is the class of Autoregressive Moving Average (ARMA) processes.

- **Non-Stationary time series**

Many applied time series, particularly arising from economic and business areas are non-stationary. Non-stationary time series can occur in many ways. They could have non-constant means, time varying variances and/or autocovariances, or all of these properties occurring simultaneously. Trend, seasonality and cyclical time series are types of non-stationary time series (Wei, 2006). One of the typical non-stationary class of models is the AutoRegressive Integrated Moving Average (ARIMA) one. A non-stationary time series can be divided in two parts: i) Homogeneous time series ii) Non-homogenous time series. In the former case, the mean is time-dependent. By computing the differences between consecutive observations, a homogeneous non-stationary time series can be converted to a stationary one. This is known as differencing. However, many non-stationary time series are non-homogenous. The non-stationarity of these series is not due to their time-dependent expected value, but rather to their time-dependent variance and autocovariance.

- **Forecasting methods**

A forecasting method is a procedure for estimating the future observations. It depends largely on what data is available. If there is no data available, or if the available data is not relevant to the forecasts, then qualitative forecasting methods must be used. There are well-developed structured approaches to obtaining good forecasts without using historical data

(Hyndman and Athanasopoulos, 2013). In contrast, quantitative forecasting can be applied when two conditions are satisfied:

1. Numerical information (data) about the past is available,
2. It is reasonable to assume that some aspects of the past patterns will continue into the future (i.e. there are no structural changes).

There is a wide range of quantitative forecasting methods, often developed within specific disciplines for specific purposes. Each method is associated with specific properties, accuracy levels and costs (of implementation) all of which must be considered when choosing between them. Most quantitative forecasting problems relate to either time series data (collected at regular intervals over time) or cross-sectional data (collected at a single point in time). Quantitative forecasting methods are divided in two general categories: 1) time series model ii) explanatory models. An explanatory model is very useful because it incorporates information about other variables, rather than only historical values of the variable to be forecast. However, there are several reasons a forecaster might select a time series model rather than an explanatory model. First, the system may not be understood, and even if it was understood it may be extremely difficult to measure the relationships assumed to govern its behaviour. Second, it is necessary to know or forecast the various predictors in order to be able to forecast the variable of interest, and this may be too difficult. Third, the main concern may be only to predict what will happen rather than explaining precisely why something happens. Finally, the time series model may give more accurate forecasts than an explanatory or mixed model (Hyndman and Athanasopoulos, 2013).

#### ▪ **Estimator selection**

In order to evaluate the impact of each aggregation approach on the forecasting performance, an estimator needs to be selected and used for extrapolation purposes. In this study, Single Exponential Smoothing (SES), also referred to as Exponentially Weighted Moving Average (EWMA) method, is used to estimate the future demand. SES is a very popular forecasting method in industry as it is intuitively appealing, easy to understand and has minimal computer storage requirements. Moreover, it is optimal for a non-stationary Integrated Moving Average process of order one, ARIMA(0,1,1). Although its application

implies a non-stationary behavior of the demand, sufficiently low smoothing constant values introduce minor deviations from the stationarity assumption whilst the method is also unbiased. SES's estimator relies upon exponentially smoothed forecasts of the demands. The estimate is updated in each period. For any time period  $t$ , the updating procedure of SES's method is presented below:

$$f_t = \alpha d_{t-1} + (1 - \alpha)f_{t-1} \quad (1-1)$$

where  $d_{t-1}$  is the demand in period  $t-1$ ,  $f_t$  is the forecast of period  $t$  and  $\alpha$  is the smoothing constant.

For any  $\alpha$  between zero and one, the weights attached to the observations decrease exponentially as we go back in time, hence the name "exponential smoothing". If  $\alpha$  is small (i.e., close to zero), the weights are spread across the observations to the very distant past. If  $\alpha$  is large (i.e., close to one), more weight is given to the more recent observations and the weights decline sharply to zero for relatively recent observations. At the extreme case where  $\alpha = 1$ , SES becomes a naïve method, i.e. the very last actual demand is the forecast for the next time period.

In this research work, we rely upon the use of the SES method rather than a popular alternative (the moving average (MA)) or any optimal forecasting method (arising under the ARIMA structure), although these forecasting methods can be considered in the next steps of research. There are two reasons that support the choice of the SES method:

i) On average, SES tends to outperform the MA method, as observed in an empirical comparison of their performance in the M3 forecasting competition (as reported by Makridakis and Hibon (2000)). In addition, SES corresponds to an intuitively appealing underlying model, whereas MA does not. It is also important to note that under the stationary assumption, Brown (1963) showed the correspondence between SES and MA (correspondence between the smoothing constant value and the length of the moving average).

ii) In practice, the decision makers may not want to spend too much time and effort examining and defining the characteristics of the data-generating process prior to determining the forecasting model, as is required by ARIMA. Besides, in a production planning framework, forecasts are required on a periodic basis, sometimes as often as on a daily or even hourly basis. Typically, forecasting is done simultaneously for several different, but related items in computerized systems with minimal human intervention. Therefore, it is quite impractical to determine the optimal ARIMA model for each item in each updating period. However, it is useful to determine the amount of gain or loss by using an optimal forecasting method instead of SES. This issue will be considered in the next steps of research.

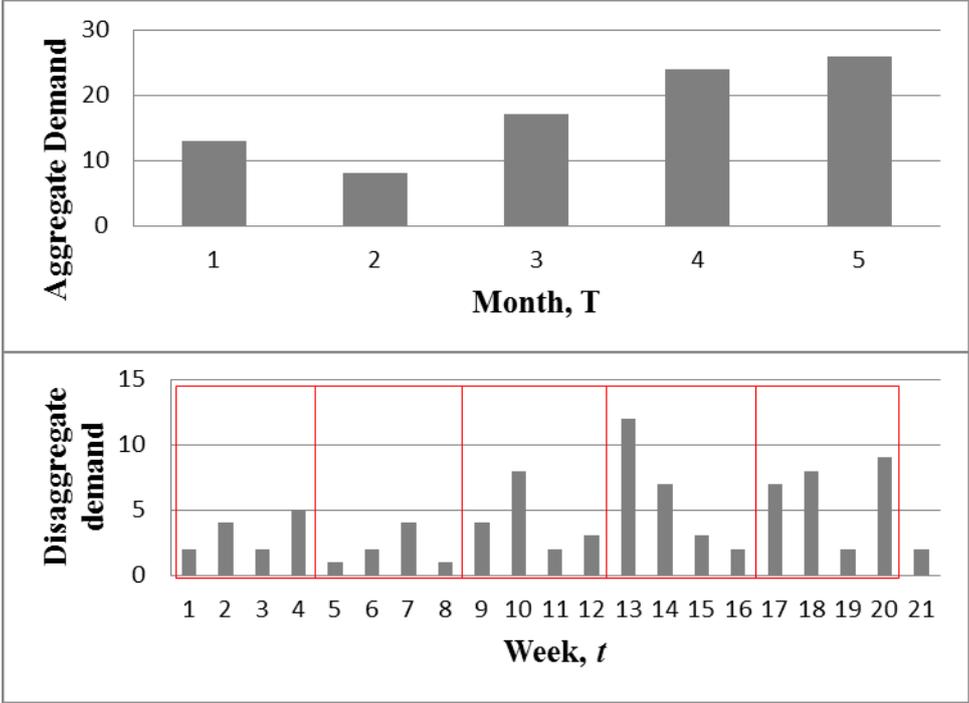
- **Accuracy measure**

An accuracy measure is a measure applied for judging the efficiency of a forecasting method. Forecast accuracy relates to a comparison between the forecast and the actual values. There are many accuracy measures discussed in the literature that may be used to report performance (Hyndman and Koehler, 2006). However, such measures are not necessarily mathematically tractable making it impossible to use them for theoretical analysis. In this research work, the variance of forecast error or equivalently the Mean Square Error (MSE) (for unbiased estimation procedures) is utilised as the only accuracy metric. Although we do wish to contrast performances on empirical data, the aim of this work is to understand the underlying reasons as to why one method performs better than another. To do so, a theoretical comparison needs to be undertaken and the MSE is the only available metric. Additionally, the *MSE* is similar to the variance of the forecast errors (which consists of the variance of the estimates produced by the forecasting method under concern and the variance of the actual demand) but not quite the same since any potential bias of the estimates may also be taken into account. Since *SES* provides unbiased estimates for the processes considered in this work the variance of forecast errors is equal to the *MSE*, *i.e.*  $MSE = Var(\text{Forecast Error})$ .

- **Demand Aggregation**

An aggregation process consists of deriving a low frequency representation of the process from a high frequency formulation; this derivation can be exerted through time or through individuals.

Aggregation across time, also called *temporal aggregation*, refers to the process by which a low frequency time series (e.g. quarterly) is derived from a high frequency time series (e.g. monthly) (Nikolopoulos et al., 2011). As shown in **Figure 1-1** and **Figure 1-2**, this is achieved through the summation (bucketing) of every  $m$  periods of the high frequency data, where  $m$  is the aggregation level. There are two different types of temporal aggregation: non-overlapping and overlapping. In the former case (**Figure 1-1**) the time series are divided into consecutive non-overlapping buckets of time where the length of the time bucket equals the aggregation level. The aggregate demand is created by summing up the values inside each bucket. The number of aggregate periods is  $[N/m]$ , where  $N$  is the number of the original periods,  $m$  the aggregation level and the  $[x]$  operator returns the integer part of  $x$ . As a consequence the number of periods in the aggregate demand is less than the original demands.



**Figure 1-1:** Non-overlapping temporal aggregation (from weekly to monthly data)

The overlapping case (**Figure 1-2**) is similar to a moving window technique where the window’s size equals to the aggregation level. At each period, the window is moved one step ahead, so the oldest observation is dropped and the newest is included. It is observed that the number of overlapping aggregate periods is higher than those of the non-overlapping and equals to  $N-m+1$ . Therefore, the information loss is negligible as compared to the non-

overlapping case. This is an important observation in terms of data availability and for the cases where little history of data is available.

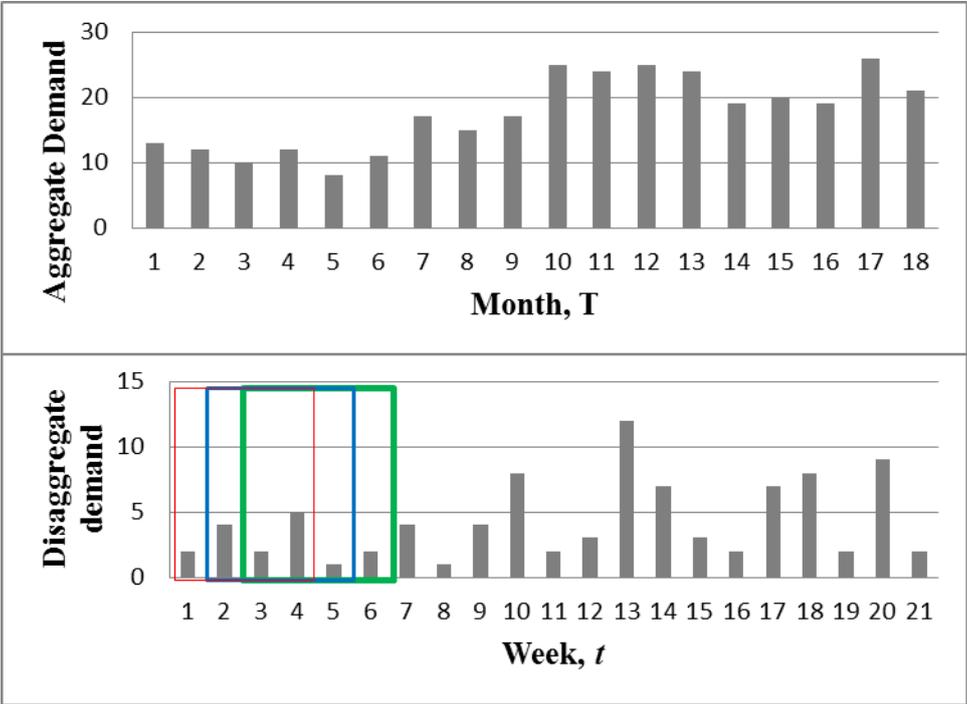


Figure 1-2: Overlapping temporal aggregation (from weekly to monthly data)

In this research, only the case of the non-overlapping temporal aggregation is considered. The overlapping temporal aggregation is an issue left for further research. In the next section, the effect of temporal aggregation on the structure of time series is reviewed.

Often, for the purpose of having comparable forecasts using the temporal aggregation approaches as compared to the classical non-aggregation approaches, if the comparison is undertaken at the disaggregate level, then the aggregate forecasts should be disaggregated to the original level (by dividing them on the aggregation level). Furthermore, if the comparison is conducted at the aggregate level, then the original forecasts should be multiplied by the aggregation level. This is illustrated in **Figure 1-3** and **1-4** in the case of weekly and monthly forecasts.

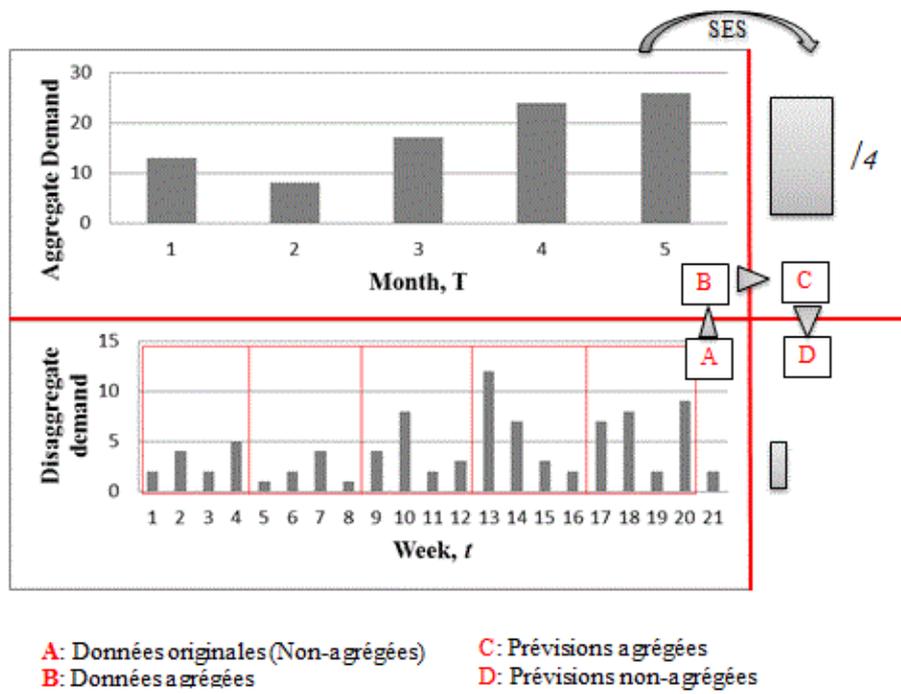


Figure 1-3: Comparison at disaggregate level

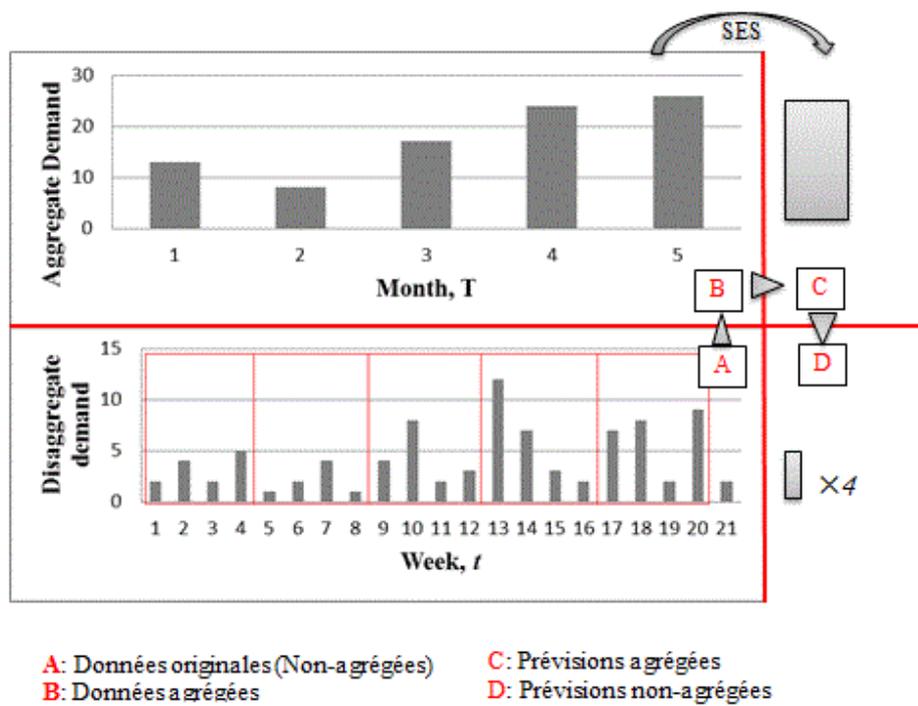
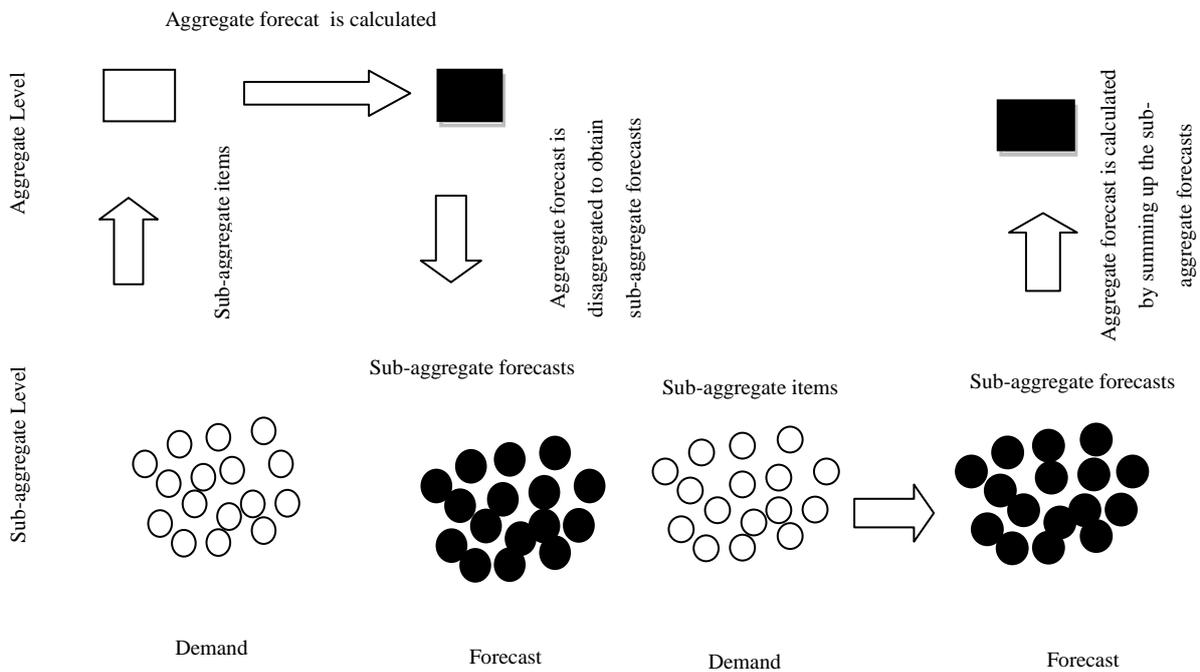


Figure 1-4: Comparison at aggregate level

Another type of aggregation referred to as *cross-sectional (or hierarchical or contemporaneous) aggregation* occurs when the aggregation takes place across a number of Stock Keeping Units (SKU) at one specific time period to reduce variability (Silvestrini and Veredas, 2008). Existing approaches to cross-sectional forecasting usually involve either a bottom-up (BU) or a top-down (TD) approach (or a combination of the two). When forecasting at the aggregate level is of interest, the former involves the aggregation of individual SKU forecasts to the group level whereas the latter relates to forecasting directly at the group level (i.e. first aggregate requirements and then extrapolate directly at the aggregate level).



**Figure 1-5:** Schematic diagram of TD (left) and BU (right) approaches

When the emphasis is on forecasting at the subaggregate level, then BU relates to direct extrapolation at the subaggregate level whereas TD involves the disaggregation of the forecasts produced directly at the group level. An important issue that has attracted the attention of many researchers as well as practitioners over the last few decades is the effectiveness of such cross-sectional forecasting approaches. As illustrated by **Figure 1-5** these approaches work as follows: The TD approach consists of the following steps: i) subaggregate demand items are aggregated; ii) the forecast of aggregate demand is produced

by applying SES at the aggregate level, and iii) the forecast is subaggregated back to the original level by applying an appropriate disaggregation method, if a subaggregate forecast is needed. In the BU approach: i) subaggregate demand forecasts are produced directly for the subaggregate items; ii) the aggregate forecast is obtained by combining individual forecasts for each SKU, i.e. potentially a separate forecasting model is used for each item in the product family (Zotteri et al., 2005). These approaches are presented schematically in **Figure 1-5**. The presentation style follows that adopted by Mohammadipour et al (2012).

## 2. Business Context

Demand forecasting is the starting point for most planning and control organizational activities. Moreover, one of the most important challenges facing modern companies is demand uncertainty (Chen and Blue, 2010). The existence of high variability in demand for fast moving and slow/intermittent moving items (items with a high ratio of zero observations) pose considerable difficulties in terms of forecasting and stock control. Deviations from the degree of variability accommodated by the Normal distribution often render standard forecasting and inventory theory inappropriate (Chen et al., 2000; Syntetos and Boylan, 2005; Wemmerlov and Whybark, 1984).

There are many approaches that may be used to reduce demand uncertainty and thus to improve the forecasting (and inventory control) performance of a company. An intuitively appealing such approach that is known to be effective is demand aggregation (Chen et al., 2007). One possibility is the *Temporal Aggregation*. Another aggregation approach often applied in practice is the *Cross-sectional Aggregation* (as discussed in the previous section). Such an approach is equivalent to aggregating data for one single SKU across a number of depots or stock locations. Natural, practically useful, associated forms of aggregation also involve geographical consolidation of data or aggregation across markets.

Although no empirical studies exist that document the extent to which aggregation takes place in practical settings, this is an approach that is known to be popular amongst practitioners because of its intuitive appeal. In practical terms, the benefit depends on the type of aggregation and of course the data characteristics. Cross-sectional aggregation for example usually leads to variance reduction. This is due to the fact that fluctuations in the data from

one time series may be offset by the fluctuations present in another time series (Widiarta et al., 2009). Contrary to cross-sectional aggregation, in temporal aggregation, variance is increased. However, it is shown that temporal aggregation can reduce the coefficient of variation of demand. In any case, the implied benefit coupled with the ease of implementing such approaches renders them a popular choice in industry.

Demand data may be broadly categorized as intermittent and fast. Aggregation of demand in lower-frequency ‘time buckets’ enables the reduction of the presence of zero observations in the former case or, generally, reduces uncertainty in the latter. Intermittent demand items (such as spare parts) are known to cause considerable difficulties in terms of forecasting and inventory modelling. The presence of zeroes has significant implications because of the following three reasons. First, the difficulty in capturing underlying time series characteristics and fitting standard forecasting models. Second, the difficulty in fitting standard statistical distributions, such as the Normal. Third, deviations from standard inventory modelling assumptions and formulations. These concerns collectively render the management of these items a very difficult exercise. Temporal aggregation is known to be applied widely in military settings (very sparse data), the after sales industry (service parts) etc. Recent empirical studies in this area (Babai et al., 2012; Nikolopoulos et al., 2011) have resulted in some very promising findings pointing out also the need for more theoretical analysis. Although the area of forecasting with temporal aggregation in an intermittent demand context is a very interesting one both from an academic and practitioner perspective, in this research only the most often occurring cases of fast demand items are considered. Analysis in an intermittent demand context is an important avenue for further research and this issue is discussed in more detail in the last chapter of this Ph.D. thesis.

In addition to the demand uncertainty reduction associated with the temporal aggregation approach discussed above, there is another important issue that relates to the forecast horizon that renders aggregation a very promising approach. The “forecast horizon” determines how far into the future the estimate projections must be. As a general rule, the further into the future we look, the more clouded our vision becomes and consequently long range forecasts are less accurate than short range forecasts. This is also one of the areas where the temporal aggregation may improve the forecast accuracy, because as we look further into

the future, the long term view becomes more important and the temporal aggregation approach may utilize this information more effectively than the classical approaches.

From an academic perspective the emphasis to date has been mainly on the cross-sectional aggregation. Moreover, most inventory forecasting software packages support the aggregation of data although this would typically cover cross-sectional aggregation only. The consideration of temporal aggregation has been somewhat neglected by software manufacturers and academics alike despite the potential opportunity for adding more value to real world practices. In this work, the objective is to advance the current state of knowledge in the area of demand forecasting temporal aggregation (and extend the existing theory on cross sectional aggregation).

In the above discussions, the effect of temporal aggregation on a single SKU is considered. However, in reality there are often many related time series that can be organized hierarchically and aggregated at several different levels in groups based on products, customers, geography or other features (Hyndman et al., 2011). The hierarchical level at which forecasting is performed depends on the function the forecasts are fed into. With regards to products (or SKUs) in particular, forecasting at the individual SKU level is required for inventory control whereas product family forecasts may be required for Master Production Scheduling. Forecasts across a group of items ordered from the same supplier may be required for the purpose of consolidating orders. Forecasts across the items sold to a specific large customer may determine transportation and routing decisions etc.

TD and BU forecasting approaches are extremely useful towards improving the accuracy of forecasts and plans when leveraged within an S&OP (Sales and Operations Planning) process (Lapide, 2006). The S&OP is a multi-functional process that involves managers from all departments (Sales, Customer Service, Supply Chain, Marketing, Manufacturing, Logistic, Procurement and Finance), where each department requires different levels of demand forecasts (Lapide, 2004). For example, in marketing (Dekimpe and Hanssens, 2000), forecasting of revenues by product groups and brands is needed; sales departments deal with sales forecasts by customer accounts and/or sales channels; supply chain managers request SKU level forecasts, while finance requires forecasts that are aggregate into budgetary units in terms of revenues and costs (Bozos and Nikolopoulos, 2011). In order to produce the required forecasts, demand and/or forecasts should be

aggregated and/or disaggregated to various levels. This involves the application of both TD and BU or a combination of them (Lapide, 2004, 2006).

### 3. Research Background

Aggregation has been widely discussed in the academic literature since as early as the 1950s (Quenouille, 1958). It is seen as a means to manage the demand fluctuation and reduce the degree of uncertainty. It has been shown by Theil (1954), Yehuda and Zvi (1960), and Aigner and Goldfeld (1974) that demand uncertainty can be effectively reduced through appropriate demand aggregation and forecasting. In the literature of supply chain planning and demand planning, demand aggregation is known as a “risk-pooling” approach to reduce demand fluctuation for more effective material/capacity planning (Chen and Blue, 2010). In the area of temporal aggregation, there are both theoretical and empirical investigations discussed in the literature. However, most of these contributions may be found in the Economics discipline. Amemiya and Wu (1972) evaluated the effect of non-overlapping temporal aggregation when the original series follows an autoregressive process of order  $p$ ,  $AR(p)$  process. By considering the ratio of MSE of non-aggregate and aggregate prediction (3 linear predictors were considered) at the aggregate level, they have shown that the aggregate approach outperforms the non-aggregate one. Tiao (1972) investigated the effect of non-overlapping temporal aggregation on a non-stationary process of the Integrated Moving Average  $IMA(d,q)$  form. A conditional expectation was applied to obtain one step ahead forecasts at the aggregate level based on the non-aggregate and aggregate series. Subsequently, the efficiency of the aggregate forecasts was defined as the ratio of the variance of the forecast error of the non-aggregate to the aggregate series when the aggregation level is large. It was shown that when  $d=0$  and the aggregation level is very high, then the ratio under concern equals one and the comparative benefit of using the non-aggregate forecasts is increased with  $d$ .

Few recent pieces of research have evaluated the effect of temporal aggregation on forecasting and stock control by means of empirical analysis. Nikolopoulos et al. (2011) empirically analysed the effects of temporal aggregation on forecasting intermittent demand requirements and they have proposed a methodology termed as ADIDA (Aggregate Disaggregate Intermittent Demand Approach to forecasting). It was shown that the ADIDA

methodology may indeed offer considerable improvements in terms of forecast accuracy. In addition, Babai et al. (2012) have extended the study discussed above (Nikolopoulos et al., 2011) by means of considering the inventory implications of the ADIDA framework through a periodic order-up-to-level stock control policy. The researchers concluded that a simple technique such as temporal aggregation can be as effective as complex mathematical intermittent forecasting approaches.

To the best of our knowledge, the only papers directly relevant to our work are those by Amemiya and Wu (1972) and Tiao (1972) for the AR and the MA process respectively. These works focused on characterizing the aggregate demand series in addition to evaluating the forecast performance. However, the results presented in these works remain preliminary in nature while their experimental setting may also be criticized in terms of the estimation procedures considered. In addition, no empirical results were obtained. Therefore, the lack of conditions that may determine the superiority of each approach in demand forecasting is obvious. It is not clear when the aggregation approach provides more accurate forecasts than the non-aggregation one and vice versa. Consequently, the motivation behind this part of the research study was the lack of the theoretical analysis regarding the effect of temporal aggregation on demand forecasting. In this research, analytical evaluation is applied to identify the superiority conditions of each approach. The research starts with the simple first order ARMA type process as discussed earlier in section 1.2. However, the analysis can be conducted for higher order processes and it this will be considered in the future.

In the area of cross-sectional aggregation, most of the forecasting literature has looked at the comparative performance of the TD and the BU approaches. The findings with regards to the performance of these approaches are mixed.

Some authors like Theil (1954), Grunfeld and Griliches (1960), Schwarzkopf et al. (1988), and Narasimhan et al., 1985(1985) argued that TD outperforms the BU approach. On the other hand another authors such as Orcutt et al. (1968) , Edwards and Orcutt (1969), Dunn et al. (1976), Dangerfield and Morris(1988) and Gross and Sohl (1990) found that the BU approach performs better; and finally some other authors like Barnea and Lakonishok (1980), Fliedner (1999) and Widiarta et al.(2007, 2008, 2009) take a contingent approach and analyse the conditions under which one approach produces more accurate forecasts than the other.

In this PhD thesis, the effectiveness of the BU and the TD approaches is evaluated. The research conducted by Widiarta et al. is extended to consider a more general stationary demand process  $ARIMA(1,0,1)$  and a non-stationary  $ARIMA(0,1,1)$  process. Moreover, the comparison is undertaken at both subaggregate and aggregate levels. Additionally, the superiority of each approach is examined by a real data set.

#### 4. Research Overview

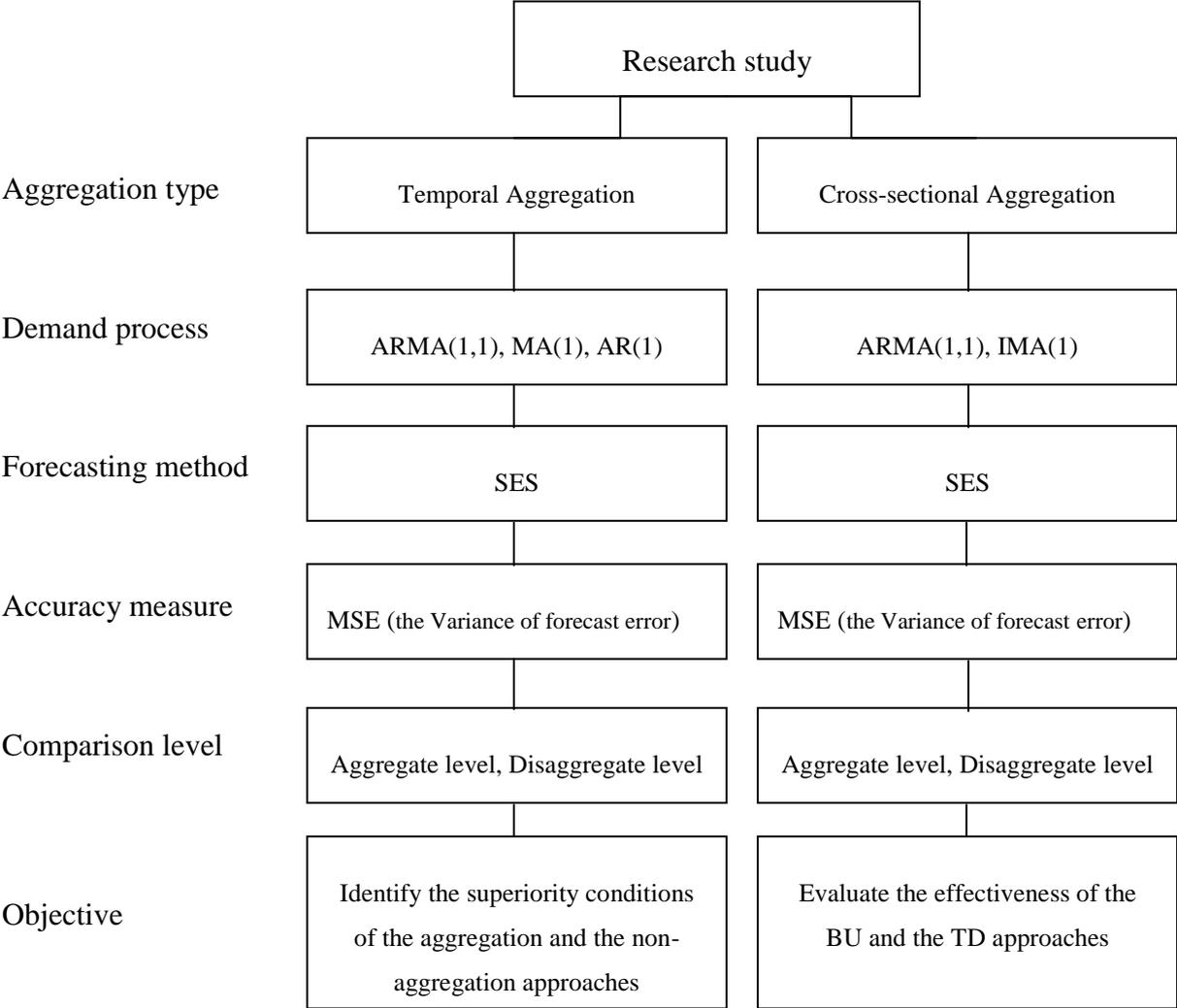
Aggregation enables forecasters to obtain forecasts at various levels across time and individual items. Depending on the level of forecasting, we may either provide the forecasts and then aggregate them or we may first aggregate the original series to obtain the aggregate demand and then produce the aggregate forecast. In the latter case, a disaggregation may be required to obtain the disaggregate forecast. In this research the impact of aggregation on demand forecasting is evaluated. To show the effect of aggregation on demand forecasting, two different types of aggregation are considered: *i*) temporal aggregation and *ii*) cross sectional aggregation. Our research overview is summarized in the **Figure 1-6**.

The mathematical analysis is complemented by a numerical investigation to validate the theoretical results which is also used in order to conduct a sensitivity analysis by some constraining assumptions considered in the analytical evaluation. Next, the findings are validated empirically (by means of simulation on a dataset provided by a European superstore) and by doing so some very much required empirical evidence in the area of demand aggregation is offered. Finally, important managerial insights are derived and tangible suggestions are offered to practitioners dealing with inventory forecasting problems.

Based on the research background and motivations, six objectives have been formulated for this research:

1. To evaluate analytically the effect of non-overlapping temporal aggregation on forecasting when the basic series follows a stationary ARMA type process.
2. To identify the conditions under which the temporal aggregation approach outperforms the non-aggregation one and vice versa.

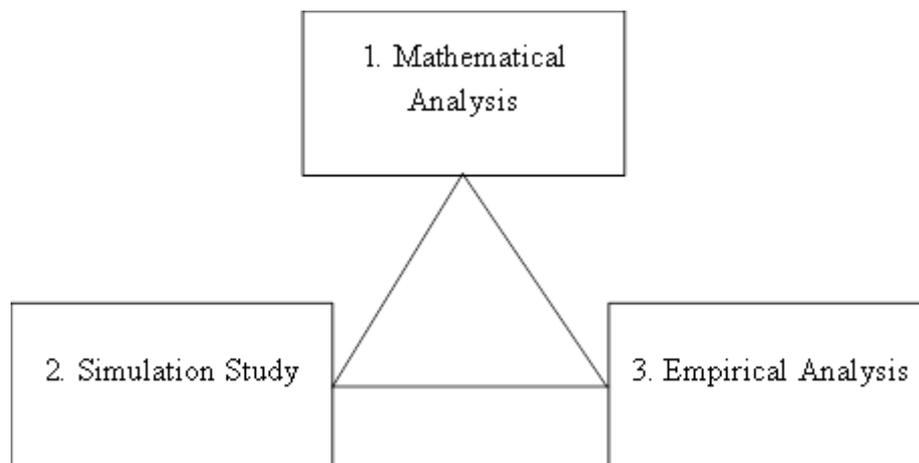
3. To determine the optimal aggregation level that maximizes the benefits of the temporal aggregation approach.
4. To examine the effectiveness of the BU and the TD approaches to forecast subaggregate and aggregate demand in a stationary and a non-stationary environment.
5. To analyse the effect of the control and the process parameters on the superiority of each approach in both temporal and cross-sectional aggregations.
6. To test the empirical validity and utility of the theoretical and simulation results on a large set of real world data.



**Figure 1-6: Research Overview**

## 5. Methodology

The research follows three research methods, namely mathematical analysis, simulation and empirical investigation. The relationship between the three methods is illustrated in **Figure 1-7**.



**Figure 1-7:** Methodology

Firstly, the mathematical analysis is applied to examine the superiority of the aggregation approach and to disclose the conditions under which this approach provides more accurate results than the classical approach. The Simulation study is used for the following reasons:

- To test and validate the results of theoretical analysis.
- To relax the assumptions considered in the mathematical evaluation.

Finally, the findings of this PhD thesis are to be tested on real empirical data to assess the practical validity and applicability of the main results of the study. Therefore, empirical analysis would help us to test the applicability of the results in real situations.

## 6. Thesis Structure

The PhD thesis is structured as follows:

In Chapter 2, an overview of demand forecasting by aggregation is presented. Different types of aggregation, i.e. temporal and cross-sectional aggregation are discussed and the effect of aggregation on process structure is described.

In Chapter 3, the effect of non-overlapping temporal aggregation on demand forecasting is examined when the underlying series follow a stationary process. For each process under consideration, the theoretical MSE is derived at both the disaggregate and the aggregate level of comparison. Then, the MSE results are compared to identify the conditions under which each approach outperforms the other. Next a simulation analysis is conducted to examine the results of the theoretical evaluation followed by an empirical investigation.

In Chapter 4, the effects of cross-sectional aggregation on demand forecasting is evaluated. It is assumed that the underlying series follow either a stationary or a non-stationary process. An analytical evaluation is first considered followed by simulation to test and validate the theoretical results. Additionally, some assumptions are relaxed compared to the theoretical analysis. The results are complemented by an empirical analysis to validate the findings on a real demand data set.

Finally, the findings from each chapter are summarized and the conclusions of this thesis are discussed in chapter 5. Managerial implications and limitations of the research are described, along with opportunities for future research.

## Chapter 2 State of the art

---

The first chapter summarized the research works conducted in this study. It outlined the research through a summary of the research background and problems, expected results, and designated methodology. This chapter aims to provide an overview of the literature on forecasting by temporal and cross-sectional aggregation.

### 1. Introduction

Demand forecasting is the starting point for most planning and control organizational activities. In general practice, accurate demand forecasts lead to efficient operations and high levels of customer service, while inaccurate forecasts inevitably lead to inefficient, high cost operations and/or poor levels of customer service. In many organizations, one of the most important actions that may be taken to improve the efficiency and the effectiveness of the decision making process is to improve the accuracy of the demand forecasts.

When developing the demand forecasting, the practitioners need to determine in which level they should produce the forecast. Forecasters need to properly identify what is the objective of the forecasting process, in terms of time bucket (i.e., forecasts are produced on a daily level, weekly or on monthly one), and set of items the demand refers to (i.e., single item or group of items). The choice of the appropriate level of forecasting depends on the decision-making process the forecast is expected to support. For instance, forecasting at the individual SKU level is required for supply chain management, while cumulative aggregate forecast may be used for budgeting or plant design. In many organizations, several managers from all departments (Sales, Customer Service, Supply Chain, Marketing, Manufacturing, Logistic, Procurement and Finance) are involved in generating forecast, where each department requires different levels of demand forecasts (Lapide, 2004).

In addition, one of the most important factors that influence the accuracy of forecasts is demand variability. Demand dispersion and uncertainty are among the most important challenges facing modern companies (Chen and Blue, 2010). These issues have been addressed in the academic literature for many years. The existence of high dispersion in demand for fast moving and slow/intermittent moving items (items with a high ratio of zero observations) pose considerable difficulties in terms of forecasting and stock control. Deviations from the degree of variability accommodated by the Normal distribution often render standard forecasting and inventory theory inappropriate (Chen et al., 2000; Syntetos and Boylan, 2005; Wemmerlov and Whybark, 1984).

There are many approaches that may be used to reduce the demand dispersion and provide the different forecast level and consequently improve the forecasting (and inventory control) performance of a company. An intuitively appealing such strategy that is known to be effective is demand aggregation (Chen et al., 2007). One approach is to aggregate demand in lower-frequency 'time buckets', thereby reducing the presence of potential zero observations (in case of intermittent demand) or generally reduce dispersion in case of fast moving demand. Such an aggregation strategy is often referred to, in the academic literature, as *Temporal Aggregation* (Nikolopoulos et al., 2011). Another aggregation strategy discussed in the literature is the *Cross-Sectional Aggregation* (also referred to as *hierarchical*), which involves aggregating different time series to obtain higher level forecasts (Silvestrini and Veredas, 2008). Existing approaches to cross-sectional forecasting usually involve either a bottom-up (BU) or a top-down (TD) approach (or a combination of the two). Although the concept of aggregation is very simple but it plays a very important role in supply chain management (Bonomo, 2003). An interesting question raised when applying aggregation to forecast demand is how exactly does that affect the demand dispersion. The relevant impact relies entirely upon the type of aggregation – cross-sectional versus temporal. Cross sectional aggregation usually leads to variance reduction. This is due to the fact that fluctuations in the data from one Stock Keeping Unit (SKU) are offset by fluctuations in the data from other SKUs (Widiarta et al., 2009). Contrary to cross sectional aggregation, in temporal aggregation, variance is increased. Schluter and Trede (Schluter and Trede, 2011) have shown that for certain types of data generation processes, both the mean and variance of the data increase through temporal aggregation. However, it is easy to show that temporal aggregation

can reduce the coefficient of variation of demand and this issue is further discussed later in our paper.

Aggregation has been widely discussed in the academic literature since as early as the 1950s (Quenouille, 1958). In a production planning framework, many researchers have focused on the effectiveness of cross-sectional aggregation and especially on the bottom-up and top-down approaches. However there are fewer studies focusing on the effects of temporal aggregation. Moreover, and although most inventory forecasting software packages support aggregation of data, this would typically cover cross-sectional aggregation only; the consideration of temporal aggregation has been neglected by software manufacturers despite the potential opportunity for adding more value to their customers.

In the following sections, the existing researches conducted in the area of temporal and cross-sectional aggregation are presented.

## **2. Temporal Aggregation**

In this section, the effect of the temporal aggregation on the process structure discussed in the literature reviewed. Then, the impact of temporal aggregation on demand forecasting discussed in the literature is presented.

### **2.2.1 Temporal aggregation identification process**

An original time series model is presented in terms of basic time unit  $t$ . Although the original form of the model can be used to produce the forecasts, however in some cases the time frequency of the observed data may not be the same as the assumed time unit  $t$ . For these cases a temporally aggregate data may be used, so it is necessary to know the effect of aggregation on model structure of the data processes. The orders of the low frequency model (i.e. monthly) from those of the high frequency model (i.e. weekly) can be determined by temporal aggregation. i.e, If the high frequency model is an ARIMA(1,0,1), what is the low frequency model? Second, once the orders are inferred, the parameters of the low frequency model is derived from the high frequency ones, rather than estimating them.

The analysis of temporal aggregation starts with the work of Amemiya and Wu (1972). It is shown that if the original variable follows a  $p^{\text{th}}$  order autoregressive process,  $ARIMA(p,0,0)$ , then the non-overlapping aggregates follow a mixed autoregressive moving average (ARIMA) model of the  $(p,0,q^*)$ . Tiao (1972) has investigated the effect of non-overlapping temporal aggregation on a non-stationary process of the Integrated Moving Average ARIMA  $(0,d,q)$  form, where  $d$  is the integrated parameter and  $q$  is the moving average parameter. It is shown that the aggregate process is of the  $ARIMA(0,d,q^*)$ . Brewer (1973) studied the effects of non-overlapping temporal aggregation on  $ARIMA(p,0,q)$  processes. It is shown that aggregating such processes results in ARMA processes with autoregressive order  $p$  and moving average order  $r$ ,  $ARMA(p,0,r)$ . The effect of the non-overlapping temporal aggregation on  $ARIMA(p,d,q)$  process is evaluated by Weiss (1984). It is seen that the temporally aggregate process is also follow an  $ARIMA(p,d,r)$  process. Wei (1979) studied the aggregation effect on univariate multiplicative seasonal time series models. It is revealed that for an  $ARIMA$  process of order  $p,d,q \times (P,D,Q)_s$ , the corresponding aggregate process is an  $ARIMA$  of order  $p,d,r \times (P,D,Q)_s$ . Brewer(1973) also presented a generalization of the results for ARMA models with exogenous variables (ARMAX models), it is shown that the temporally aggregate  $ARIMAX(p, d, q)(k)$  model is an  $ARIMAX(p, d, r)(a)$ .

Teles et al (1999) showed that temporal aggregation changes the order of a fractionally integrated ARFIMA process to an  $ARFIMA(p,d,\infty)$ , while leaving the value of  $d$  unchanged. Additionally, Souza and Smith (2004) showed that for AR Fractionally IMA (ARFIMA) models temporal aggregation results in bias reduction.

Drost and Nijman (1993) considered the effect of temporal aggregation on the ARMA models with symmetric GARCH errors,  $ARMA(p,q)$ - $GARCH(P,Q)$ . It is revealed that the aggregate model follows an  $ARMA(p,r)$  with weak  $GARCH(R,R)$ . They have also considered the ARCH and GARCH type models. It is shown that the temporal aggregation of an  $ARCH(q)$  is an  $GARCH(q,q)$ , it is also seen that the temporally aggregate  $GARCH(1,q)$  is an  $GARCH(q,q)$ .

Stram and Wei (1986) studied the relationship between the autocovariance function of disaggregate and aggregate processes. They have shown that the autocovariance function of the latter can be computed based on the autocovariance function of former; in particular the

autocovariance function after aggregation is a function of the aggregation level and autocovariance function before aggregation.

**Table 2-1** summarized the effect of the non-overlapping temporal aggregation on the structure of the process.

**Table 2-1:** the effect of the non-overlapping temporal aggregation on process's structure

Non-aggregate process	Aggregate process	Parameters	Reference
ARIMA $(p, 0, q)$	ARMA $(p, q^*)$	$q^* = \left[ p + 1 + \frac{q - p - 1}{m} \right]$	(Brewer, 1973)
ARIMA $(p, 0, q)$	ARMA $(p, q^*)$	$q^* = \left[ \frac{(m-1)(p+1)}{m} \right]$	(Amemiya and Wu, 1972)
ARIMA $(0, 0, q)$	MA $(n_0)$	$n_0 \leq q^* = \left[ 1 + \frac{q-1}{m} \right]$	(Wei, 2006)
ARIMA $(0, d, q)$	IMA $(d, n_0)$	$n_0 \leq q^* = \left[ d + 1 + \frac{q-d-1}{m} \right]$	(Tiao, 1972)
ARIMA $(p, d, q)$	ARIMA $(p, d, r)$	$r = \left[ \frac{p(m-1) + (d+1)(m-1) + q}{m} \right]$	(Weiss, 1984)
ARIMA $(p, d, q)$ $(P, D, Q)_s$	ARIMA $(p, d, r)$ $(P, D, R)_s$	$r = \left[ \frac{(p+1)(m-1) + d(m-1) + q}{m} \right]$	(Wei, 1979)

$$R = \left[ \frac{(P + D)s * k + (Q - P - D)s}{m} \right]$$

First, if  $m < s$ , there is still some seasonality in the temporally aggregate process. Second, if  $m$  is a multiple of  $s$ , the seasonal cycle remains constant. Last, if  $m$  is equal or larger than  $s$ , seasonality vanishes.

		$r > \left[ \frac{(p + d + 1)(k - 1) + q}{k} \right]$	(Brewer, 1973)
ARIMAX( $p, d, q$ )( $k$ )	ARIMAX( $p, d, r$ )( $a$ )	$r > \left[ \frac{(p + d + 1)(k - 1) + m - 1 + d}{k} \right] + \left[ \frac{(v + d)(k - 1) + \omega}{k} \right]$ $a = \left[ \frac{(p + d + 1)(k - 1) + m - 1 + d}{k} \right] + v + d$	
ARFIMA( $p, d, q$ )	ARFIMA( $p, d, \infty$ )	-	(Teles et al., 1999)
ARCH( $q$ )	GARCH( $q, q$ )	-	(Drost and Nijman, 1993)
GARCH(1, $q$ )	GARCH( $q, q$ )	-	(Drost and Nijman, 1993)

$$r = \left[ \frac{(p+1)(m-1) + q}{m} \right]$$

(Drost and Nijman, 1993)

ARIMA( $p, 0, q$ )-  
GARCH( $P, Q$ )

ARMA( $p, r$ ) with weak  
GARCH( $R, R$ )

$$R = \bar{r} + \frac{1}{2} r(r+1)$$

$$\bar{r} = \max(P, Q)$$

INAR(1)

INARMA(1,1)

-

(Brannas et al., 2002)

INMA(1)

INMA(1)

-

(Brannas et al., 2002)

Luiz et al.(1992) evaluated the effect of overlapping temporal aggregation where the original series follows an ARIMA process. It is found that the temporally aggregate process of an ARIMA(p,d,q) is an ARIMA(P,d,Q). To the best of our knowledge this is the only research dealing with the impact of overlapping temporal aggregation on the structure of the ARIMA type process.

**Table 2-2:** the effect of the overlapping temporal aggregation on process's structure

ARIMA(p,d,q)	ARIMA(P,d,Q)	$P \leq p$ and $Q \leq q+m-1$	(Luiz et al., 1992)
--------------	--------------	-------------------------------	---------------------

Although many studies consider the case of fast moving items or continuous-valued time series, integer time series have received less attention in a temporal aggregation context. Brannas et al(2002) first studied the non-overlapping temporal aggregation of an Integer Auto-Regressive process of order one, INARIMA(1,0,0), It is shown that the aggregate series follows an Integer Auto-Regressive Moving Average process of order one, INARIMA (1,0,1). Additionally, it is observed that the non-overlapping temporal aggregation of an Integer moving average process of order one, INARIMA(0,0,1) is an INARIMA(0,0,1).

**Table 2-3:** the effect of temporal aggregation on integer ARIMA type process's structure

Non-aggregate process	Aggregate process	Type	Reference
INARMA (p,0, q)	INARMA (p,0, q)	Overlapping	(Mohammadipour and Boylan, 2012)
INARMA(1,0, 0)	INARMA(1,0, 0)	Overlapping	(Brannas et al., 2002)
INARMA(1,0, 0)	INARMA(1,0, 1)	Non-overlapping	(Brannas et al., 2002)
INARMA(0,0, 1)	INARMA(0,0, 1)	Non-overlapping	(Brannas et al., 2002)

Brannas et al.(2002) evaluated the effect of overlapping temporal aggregation for the INARIMA(1,0,0) process, it is seen that the aggregate process also follow an INARIMA(1,0,0) process. The effect of overlapping temporal aggregation on INARIMA( $p,0,q$ ) process is evaluated by Mohammadipour and Boylan (2012). It is shown that the overlapping temporally aggregate of an INARIMA( $p,0,q$ ) process is also an INARIMA( $p,0,q$ ) one.

In the next section we provide a review of the studies that apply temporal aggregation approach in the area of demand forecasting.

## 2.2.2 Demand forecasting by temporal aggregation

In the supply chain and demand planning literature, demand aggregation is generally known as a 'risk-pooling' approach to reduce demand fluctuation for more effective material/capacity planning (Chen and Blue, 2010). Demand uncertainty may considerably affect forecasting performance with further detrimental effects in production planning and inventory control. It has been shown by Theil (Theil, 1954), Yehuda and Zvi (Yehuda and Zvi, 1960), Aigner and Goldfeld (Aigner and Goldfeld, 1974) that demand uncertainty can be effectively reduced through appropriate demand aggregation and forecasting.

Most of the literature that deals with temporal aggregation may be found in the Economics discipline. The analysis of temporal aggregation starts with the work of Amemiya and Wu (Amemiya and Wu, 1972). They assumed that the original variable follows a  $p^{th}$  order autoregressive process,  $AR(p)$ . By considering the ratio of MSE of disaggregate and aggregate prediction (3 linear predictors were considered) at the aggregate level, they have shown that the MSE of disaggregate forecasts is greater than that of the aggregate ones, i.e. the aggregation approach outperforms the non-aggregation one. Tiao (Tiao, 1972) has investigated the effect of non-overlapping temporal aggregation on a non-stationary process of the Integrated Moving Average IMA( $d,q$ ) form. They applied a conditional expectation to obtain one step ahead forecasts at the aggregate level based on the disaggregate and aggregate series. Subsequently, the efficiency of the aggregate forecasts was defined as the ratio of the variance of the forecast error of the disaggregate to the aggregate series when the aggregation

level is large. They have shown that when  $d=0$  the ratio under concern equals to 1 and the comparative benefit of using the disaggregate forecasts is increasing with  $d$ .

Athanasopoulos et al. (2011) have looked at the effects of non-overlapping temporal aggregation on forecasting accuracy in the tourism industry. They conducted an empirical investigation using 366 monthly series and some forecasting methods tested in the M3 competition data (Makridakis and Hibon, 2000), namely Innovations state space models for exponential smoothing (labeled ETS), the ARIMA methodology, a commercial software (Forecast Pro), damped trend (Gardner and McKenzie, 1985), the Theta method and naïve. The monthly series were aggregate to be quarterly, and the quarterly series were further aggregate to be yearly. Subsequently, they compared the accuracy of the forecasts made before and after aggregation. They considered one and two step-ahead forecasts and three statistical measures were used to compare the results: Mean Absolute Percentage Error (*MAPE*), Mean Absolute Scaled Error (*MASE*) and Median Absolute Scaled Error (*MdASE*). The aggregate forecasts at the yearly level (whether produced from monthly or quarterly data) were found to be more accurate than the forecasts produced from the yearly data directly. This study provided considerable empirical evidence in support of temporal aggregation.

Luna and Ballini (2011) have used a non-overlapping aggregation approach to predict daily time series of cash money withdrawals in the neural forecasting competition, NN5<sup>2</sup>. Each time series consisted of 735 daily observations which have been used to forecast 56 daily steps ahead for two sets of 11 and 111 time series. Daily samples were aggregate to give weekly time series and then an adaptive fuzzy rule-based system was applied to provide 8-step-ahead forecasts (thus aggregation reduced the forecast horizon from 56 to 8 steps). Two different aggregation approaches were evaluated for this purpose: the historical top-down (TD-H) approach and the daily top-down (TD-DM) approach, where the main difference between the two was the disaggregation procedure. In the former case aggregate forecasts were dis-aggregate based on historical percentages. In the latter case, the daily estimations were ‘corrected’ by multiplying them by the associated weekly estimation and dividing by the sum of the seven daily estimated samples. The symmetric MAPE (*sMAPE*) and the Mean

---

<sup>2</sup> <http://www.neural-forecasting-competition.com/NN5/>

Absolute Error (*MAE*) were used to compare the results. The researchers showed that the aggregate forecasts produced by the two approaches performed similarly or better than those given by the daily models directly. The reduction of a forecast horizon from 56 to 8 steps ahead would be intuitively expected to lead to performance improvement.

The effect of temporal aggregation on demand forecasting for integer time series have received less attention comparing to continues time series. Willemain et al.(1994) empirically explored the effects of temporal aggregation on forecasting intermittent demand considering the application of Croston's method(Croston, 1972) that has been specifically developed for such demand patterns. The researchers considered 16 empirical data sets of 905 daily observations; the aggregation level was considered to be a week. Results were reported by considering the *MAPE* and the researchers showed a significant reduction in forecasting errors when weekly demand aggregate data were used instead of daily data.

Mohammadipour and Boylan (2012) have studied theoretically the effects of overlapping temporal aggregation of INARMA processes. They showed that the aggregation of an INARMA process over a given horizon results in an INARMA process as well. The conditional mean of the aggregate process was derived as a basis for forecasting. A simulation experiment was conducted to assess the accuracy of the forecasts produced using the conditional mean of the aggregation approach for three INARMA processes: INARIMA(1,0,0), INARIMA(0,0,1) and INARIMA(1,0,1), against that of the non-aggregation approach. The simulation results showed that, in most cases, the aggregation approach provides forecasts with smaller MSEs than non-aggregation ones. The performance of these forecasts was also tested by using two empirical datasets. The first one was from the Royal Air Force (RAF, UK) and consisted of the individual demand histories of 16,000 SKUs over a period of 6 years (monthly observations). The second data set consisted of the demand history of 3,000 SKUs from the automotive industry (over a period of 24 months). The outcome of the empirical investigation confirmed the simulation results.

Nikolopoulos et al. (2011) have empirically analysed the effects of non-overlapping temporal aggregation on forecasting intermittent demand requirements. Their proposed approach, called Aggregate-Disaggregate Intermittent Demand Approach (ADIDA), was assessed on 5,000 SKUs containing 7 years history (84 monthly demand observations) form the Royal Air Force (RAF, UK), by means of employing three methods: Naïve, Croston and

Syntetos-Boylan Approximation (SBA)(Syntetos and Boylan, 2005). The aggregation level was varied from 2 to 24 months. Comparisons were performed at the original series level (disaggregate demand) and the results showed that the proposed ADIDA methodology may indeed offer considerable improvements in terms of forecast accuracy. The main conclusions of this study were: (1) the ADIDA may be perceived as an important method *self-improving mechanism*; (2) an optimal aggregation level may exist either at the individual series level or across series; (3) setting the aggregation level equal to the lead time length plus one review period  $L+1$  (which is the time bucket required for periodic stock control applications) shows very promising results. Spithourakis et al.(Spithourakis Georgios P. et al., 2011) extended the application of the ADIDA approach to fast-moving demand data. The method's performance was tested on 1,428 monthly time series of the M3-Competition by using the Naïve, SES, Theta, Holt and damped forecasting methods. The empirical results confirmed the previous findings reported by Nikolopoulos et al.(Nikolopoulos et al., 2011).

Finally, Babai et al. (2012) have also extended the study discussed above (Nikolopoulos et al., 2011) by means of considering the inventory implications of the ADIDA framework through a periodic order-up-to-level stock control policy. Three forecasting methods, SES, Croston and SBA were used and the demand was assumed to be negative binomially distributed. Performance was reported through the inventory holding and backlog volumes and costs, for three possible targets Cycle Service Levels (CSL): 90%, 95% and 99%. For high CSLs, the aggregation approach has been shown to be more efficient but for low CSLs it was outperformed by the classical one when Croston's method was used. For SES, the aggregation approach outperforms the classical approach even for low CSLs. The researchers concluded that a simple technique such as temporal aggregation can be as effective as complex mathematical intermittent forecasting approaches.

### **3. Cross-sectional aggregation**

In this section, the effect of cross-sectional aggregation on the process structure is summarized. Then, the effectiveness of cross-sectional aggregation approaches on demand forecasting in the literature is reviewed.

### 2.3.1 Cross-sectional aggregation identification process

When dealing with the impact of the cross-sectional aggregation on forecasting, it's necessary to infer the characteristics of the aggregate data from the original subaggregate data. i.e, If the subaggregate series follow an ARIMA process, it is possible to investigate whether the aggregate observed series follows an ARIMA process as well.

(Granger and Morris, 1976) showed that the cross-sectional aggregation of  $N$  uncorrelated ARIMA( $p,0,q$ ) processes is also an ARIMA( $x,0,y$ ) process. As a special case the showed that the sum of two uncorrelated ARMA processes, ARMA ( $p_1, q_1$ ) and ARMA ( $p_2, q_2$ ) is also an ARMA ( $p_1 + p_2, K$ ), where  $K \leq \max(p_1 + q_2, p_2 + q_1)$ .

Anderson (1975) stated that the sum of  $N$  independent Moving Average processes: MA( $q_1$ ), MA( $q_2$ )...MA( $q_n$ ), is an MA ( $q$ ) process as well. It is seen by Harvey (1993) that when the subaggregate items follow and ARIMA(1,0,0) process, the aggregate data may follow an ARIMA(1,0,0), ARIMA(2,0,0) or ARIMA(2,0,1) process. Zaffaroni (2007) showed that the sum of two independent strong GARCH(1,1) processes is weak GARCH(2,2).

**Table 2-4:** aggregate process of cross-sectional aggregation

Sub-aggregate process	Aggregate process	Parameter	Reference
ARIMA( $p_i, 0, q_i$ )	ARIMA( $x, 0, y$ )	$x \leq \sum_{i=1}^N p_i$ $y \leq \max(x - p_i + q_i)$	(Granger and Morris, 1976)
AR(1)+ AR(1)	AR(1)	If $\phi_1 = \phi_2$	(Harvey, 1993)
	AR(2)	If $\phi_1 = -\phi_2$	
	ARMA(2,1)	otherwise	
MA( $q_i$ )	MA( $q$ )	$q \leq \max(q_1, q_2, \dots, q_n)$	(Anderson, 1975)
GARCH(1,1)	GARCH(2,2)	When $\alpha_1 + \beta_1 = \alpha_2 + \beta_2$	(Zaffaroni, 2007)

### 2.3.2 Demand forecasting by cross-sectional aggregation

Demand forecasting for sales and operations management often concerns many items, perhaps hundreds of thousands, simultaneously. The conventional forecasting approach is to extrapolate the data series for each SKU individually. However, most businesses have natural groupings of SKUs; that is, the SKUs may be aggregate to get higher levels of forecasts across different dimensions such as product families, geographical area, customer type, supplier type etc (Chen and Boylan, 2007). Such an approach enables the potential identification of time series components such as trend or seasonality that may be hidden or not particularly prevalent at the individual SKU level. Group approaches for example are known to offer considerable benefits towards the estimation of seasonal indices (Chen and Boylan 2008). Most of the forecasting literature in this area has looked at the comparative performance of the top-down (TD) and the bottom-up (BU) approach. The findings with regards to the performance of these approaches are mixed.

Many researchers have provided evidence in favour of the TD approach. Gross and Sohl (1990) for example, numerically found that the TD approach (in conjunction with an appropriate disaggregation method) provided better estimates than BU forecasting in two out of three product lines examined. Fliedner (1999) evaluated by means of simulation the forecast system performance at the aggregate level resulting from varying degrees of cross correlation between two subaggregate time series. The subaggregate items were assumed to follow a Moving Average process of order one, MA(1) and the forecasting methods considered were SES and the Simple Moving Average (SMA). This research showed the forecast performance at the aggregate level to benefit from the TD approach. Barnea and Lakonishok (1980) examined the effectiveness of BU and TD on forecasting corporate performance. They reported that positive cross-correlation contributes to the superiority of forecasts based on aggregate data (TD).

On the other hand, Orcutt et al. (1968) and Edwards and Orcutt (1969) argued that information loss is substantial when aggregating and therefore the bottom-up approach provides more accurate forecasts. Dangerfield and Morris (1992) and Gordon et al. (1997)

used a subset of the M-competition<sup>3</sup> data (Makridakis et al., 1982) to examine the performance of TD and BU approaches on subaggregate demand forecasting. They found that forecasts by the BU approach were more accurate in most situations especially when items were highly correlated or when one item dominated the aggregate series. Weatherford et al. (2001) evaluated the performance of BU and TD approaches to obtain the required forecasts for hotel revenue management. The data they considered was perceived as very typical within the hotel industry. They experimented with four different approaches (fully subaggregate d, aggregating by rate category only, aggregating by length of stay only, and aggregating by both rate category and length of stay) to get detailed forecasts by day of arrival, length of stay and rate category and length of stay for revenue management. The results of their study showed that a purely subaggregate forecast strongly outperformed even the best aggregate forecast.

Some authors take a contingent approach and analyse the conditions under which one method produces more accurate forecasts than the other. Shlifer and Wolff (1979) evaluated analytically the superiority of BU and TD on forecasting sales for specific and entire market segments. They mentioned that BU is preferable for the purpose of forecasting the aggregate series. In addition, they found that increasing the number of SKUs favours TD. However, when the comparison was performed at the subaggregate level, they found that TD often results in larger forecast error than BU. Lütkepohl (1984) showed that it might be preferable to forecast aggregate variables using a TD approach when a time series is generated by a multivariate ARMA process and the statistical properties of the subaggregate items are known. However, if the processes used for forecasting are estimated from a given set of time series data then the BU approach outperformed TD. Widiarta et al. (2007) studied analytically the conditions under which one approach outperforms the other for forecasting the item level demands when the subaggregate items follow a first-order autoregressive [AR(1)] process with the same autoregressive parameter for all the items and when SES is used to extrapolate future demand requirements. They found that the superiority of each approach is a function of the autoregressive parameter. Widiarta et al. (2008, 2009) also evaluated analytically the effectiveness of TD and BU approaches at the subaggregate and aggregate level, respectively.

---

<sup>3</sup> The M Competition is an empirical forecast accuracy comparison exercise introduced by Prof. Makridakis.

They showed that when all subaggregate items follow an MA(1) process with identical moving average parameters, there is no difference in the relative performance of TD and BU forecasting as long as the optimal smoothing constant is used in both approaches. Subsequently, they conducted a simulation analysis considering non-identical process parameters for subaggregate items and concluded that there is significant difference between the two approaches. The superiority of each approach was a function of the moving average parameter, the cross-correlation and the proportion of a subaggregate component's contribution to the aggregate demand. Viswanathan et al. (2008) used a simulation study to investigate the effectiveness of TD and BU approaches in estimating the aggregate data series when the subaggregate items are intermittent. The study reveals that low variability of the inter-demand intervals favours the BU approach (using Croston's method (Croston, 1972)). However, when demand sizes and inter-demand intervals of the subaggregate series are highly variable and aggregation encompasses many items, TD performs best.

#### **4. Discussion on the literature review**

In this chapter, an overview of the literature on the demand forecasting by aggregation approach is given. The overview presented by classifying the literature into two parts: temporal and cross-sectional aggregation approaches.

In the first part, the theoretical and empirical investigations in the area of temporal aggregation are discussed. The former mainly focused on the structure of the aggregate time series and the relationship between the aggregate and disaggregate process parameters. The latter evaluated the effect of the temporal aggregation on demand forecasting in terms of forecast accuracy measures and stock control metrics. According to the literature, temporal aggregation approach may provide more accurate forecasts than classical one in the fast and slow moving environments. However, the conditions under which one approach may outperform other one are not discussed in the literature. It is not clear when disaggregate data should be used and where it is better to use the aggregate data to produce the forecast.

To the best of our knowledge, the only papers directly relevant to our work are those by Amemiya and Wu (1972) and Tiao (1972) for the AR and MA process respectively. In both cases the researchers investigated the forecast performance of temporal aggregation

strategies under an (Auto-Regressive Integrated Moving Average) ARIMA-type framework. However, the results presented in these two papers remain preliminary in nature while the experimental setting may also be criticized in terms of the estimation procedures considered. In addition, no empirical results were provided. Important as they are, both papers focused on characterizing the aggregate demand series rather than the forecast performance.

In this research work, the conditions under which aggregation and non-aggregation approaches yield more accurate forecasts are determined by analytical investigation. This work considers the case of ARIMA(1,0,1) and its special cases ARIMA(0,0,1) and ARIMA(1,0,0) processes and as such some of the theoretical results presented in the above discussed research are of direct relevance to our analysis. Our work differs from these works though and extends them in some very significant ways: i) optimal estimators are seldom used in practice not only due to the computational requirements that are typically prohibitive but also the lack of understanding on the part of the managers of their functionality. In addition, there is evidence to support the fact that simple forecasting methods (such as SES that is used in our work) perform at least as good as more complex theoretically coherent alternatives (Makridakis and Hibon, 2000); ii) a difficulty associated with aggregation methods is the fact that a disaggregation mechanism is also required since very often forecasts are needed at the original/disaggregate demand level. Both papers consider a comparison at the aggregate level which addresses only part of the forecasting problem. Consideration of a comparison at the original demand level, which is the case considered in this work, addresses another part of the problem and is an important extension of the research already being done<sup>4</sup>; iii) no empirical analysis has been undertaken in both papers in contrast with this work where the theoretical findings are empirically validated; iv) the analysis is complemented by means of further

---

<sup>4</sup> An important assumption in our analysis is that we start with data that are as disaggregate as our required forecasting output. However, and as one of the referees correctly pointed out the degree of aggregation of the forecasting output does not necessarily need to match with the existing data structure (which may be more aggregate or more disaggregate than the forecasts driving decision making). The degree of aggregation of the forecasting output (i.e. the forecast we use to make decisions) is actually a function of the decision making problem forecasting tries to support. On the contrary inputs to the forecasting process are very often driven by existing data structures. Although the two may indeed match sometimes, this is not always the case.

numerical investigations to identify the optimum aggregation level and smoothing constant values that require to be used.

In the second part, the comparative performance of the BU and TD approaches to forecast subaggregate and aggregate demand is reviewed. Most of the researches in this area is based on the simulation and the empirical analysis. However, there are few work focused on the effectiveness of BU and TD by analytical investigation.

To the best of our knowledge, the only papers directly relevant to our work are those by Widiarta et al.(2007, 2009) and Sbrana and Silvestrini (2013). Widiarta et al. evaluated analytically the effectiveness of the TD and BU approaches under the assumption of an AR(1) comparing at subaggregate level(Auto-Regressive process of order 1) and MA(1) process comparing at aggregate level respectively. Sbrana and Silvestrini identify the condition of superiority of Bu and TD compared at aggregate level when the demand process follow and ARIMA(0,1,1) process with non identical parameters.

In summary what can be concluded from the cross-sectional literature is both BU and TD approaches appear to be associated with superior performance. This superiority depends on the structure of the series and cross-correlation related assumptions.

In this work the relative effectiveness of the BU and TD approach for forecasting is evaluated. It is recognized that forecasts may be equally required at both the aggregate and sub-aggregate level and as such comparisons are performed at both levels. In addition, a more general univariate stationary and a non-stationary demand processes at both aggregate and subaggregate levels are studied. Moreover, the analysis is complemented by means of an empirical investigation using real data.

## Chapter 3 Temporal aggregation

---

In chapter 2, an overview of the literature on demand forecasting by aggregation is provided, additionally the necessity of conducting more research work in the area of demand aggregation is discussed. In this chapter the effect of temporal aggregation on demand forecasting by means of the analytical, simulation and empirical investigation is evaluated. The conditions under which temporal aggregation may improve the accuracy of the demand forecasts are identified. The effects of temporal aggregation on forecasting when the underlying series follows a first order Autoregressive Moving Average process,  $ARIMA(1,0,1)$  Autoregressive process of order one,  $ARIMA(1,0,0)$  and a Moving Average process of order one,  $ARIMA(0,0,1)$  is studied. Furthermore, the forecasting method is the Single Exponential Smoothing (SES). These assumptions bear a significant degree of realism. As it is discussed later in the chapter there is evidence to support the fact that demand often follows the stationary processes assumed in this work (48% of the empirical series available in our research follow such processes). Moreover, SES is a very popular forecasting method in the industry (Acar and Gardner, 2012; Gardner, 1990, 2006; Taylor, 2003). Although its application implies a non-stationary behavior of the demand, sufficiently low smoothing constant values introduce minor deviations from the stationarity assumption whilst the method is also unbiased.

In this chapter the variance of the forecast error (or equivalently, by considering an unbiased estimation procedure, the mean square error) obtained based on the aggregate demand to that of the non-aggregate demand is analytically compared. Comparisons are performed at both disaggregate and aggregate demand level. It is mathematically shown that the ratio of the Mean Squared Error (MSE) of the latter approach to that of the former is a function of the aggregation level, the process parameters and the exponential smoothing constant. The mathematical analysis is complemented by a numerical investigation to test and validate the results. Next, the theoretical results are validated empirically (by means of simulation on a

dataset provided by a European superstore) and by doing so some very much needed empirical evidence in the area of temporal aggregation are offered.

To the best of our knowledge, the only works directly relevant to our work are those by Amemiya and Wu (1972) and Tiao (1972) for the AR and MA process respectively. In both cases the researchers investigated the forecast performance of temporal aggregation strategies under an (Auto-Regressive Integrated Moving Average) ARIMA-type framework. However, the results presented in these two papers remain preliminary in nature while the experimental setting may also be criticized in terms of the estimation procedures considered (Zotteri and Kalchschmidt, 2007). In addition, no empirical results were provided. Important as they are, both works focused on characterizing the aggregate demand series rather than the forecast performance as explained in the chapter 2.

This study attempts to fill this gap and provides helpful guidelines to select the appropriate approach under such demand processes. The work discussed in this chapter can be extended to analyse more general cases such as  $ARIMA(p,0,0)$ ,  $ARIMA(0,0,q)$  or even  $ARIMA(p,0,q)$  processes. However, the analysis and presentation of such results would become too complex. Since the main objective of this research is to obtain some key managerial insights, the analysis is restricted to the  $ARIMA(1,0,1)$ ,  $ARIMA(1,0,0)$  and  $ARIMA(0,0,1)$  processes only.

Considerable part of this chapter has been published in Rostami-Tabar et al (2013a) and Rostami-Tabar et al (2013c).

This chapter is organized as follows. In section 1, theoretical analysis of temporal aggregation for autoregressive moving average process order one,  $ARIMA(1,0,1)$  and its special cases moving average order one,  $ARIMA(0,0,1)$  and autoregressive order one,  $ARIMA(1,0,0)$  is evaluated. In section 2 the results of the theoretical evaluation obtained in sub-section 1 is presented. In section 3 the simulation investigation to test and validate the results of the mathematical analysis is used. Next, a real data set to validate the results of theoretical and simulation parts in practice is applied in section 4. Finally the conclusions are given in section 5.

## 1. Theoretical Analysis

In this section the variance of the forecast errors generated by considering the disaggregate and the aggregate demand is derived. Comparisons are performed at the original disaggregate and aggregate level. To that end, the aggregation approach works as follows: firstly buckets of aggregate demand are created based on the aggregation level. Then *SES* is applied to this aggregate data to produce the aggregate forecasts, now if the comparison is undertaken at aggregate level then the aggregate forecast is maintained, however, to compare at disaggregate level the aggregate forecasts are disaggregated by dividing by  $m$  to produce forecasts at the original level. In addition other disaggregation mechanisms could have been considered (Nikolopoulos et al., 2011) but the one employed for the purposes of this research is viewed as realistic from a practitioner's perspective and seen as a reasonable approach when dealing with stationary demands. Note that in order to ensure that the forecasting horizon is the same in both the aggregate and the disaggregate cases, the aggregate *SES* forecast is updated in each period when the aggregate series are rebuilt.

The comparisons result in the development of theoretical rules that indicate under which conditions the forecasting of the aggregate demand is theoretically expected to perform better than the forecasting of the disaggregate demand. These theoretical rules are a function of the aggregation level, the control, and the process parameters. The cut-off values to be assigned to the parameters are the outcome of a numerical analysis to be conducted based on the theoretical results. Having obtained the cut-off values, we can then specify regions of superior performance of the aggregation approach over the non-aggregation one.

In this study the variance of the forecast error is used as a forecast accuracy measure as it is the only theoretically tractable measure. The *MSE* is similar to the variance of the forecast errors (which consist of the variance of the estimates produced by the forecasting method under concern and the variance of the actual demand) but not quite the same since any potential bias of the estimates may also be taken into account (Syntetos., 2001). Since *SES*

provides unbiased estimates<sup>5</sup> (due to the stationarity of the time series considered in this work) the variance of forecast errors is equal to the *MSE*, i.e.  $MSE = Var(\text{Forecast Error})$ .

For each process under consideration the ratio of the MSE before aggregation ( $MSE_{BA}$ ) to the MSE after aggregation ( $MSE_{AA}$ ) is calculated. A ratio that is lower than one implies that the aggregation approach does not add any value. Conversely, if the ratio is greater than one, aggregation approach performs better than the classical one.

### 3.1.1 Notation and assumptions

For the remainder of the research the notations are denoted by:

$m$ : Aggregation level, i.e. number of periods considered to build the block of aggregate demand.

$n$ : total number of periods available in the demand history.

$t$ : Time unit in the original disaggregate time series.  $t=1,2,\dots,n$ .

$T$ : Time unit in the aggregate time series.  $T=1,2,\dots, \lceil n/m \rceil$ .

$d_t$ : Disaggregate demand in period  $t$

$D_T$ : Aggregate demand in period  $T$

$\varepsilon_t$ : Independent random variables for disaggregate demand in period  $t$ , normally distributed with zero mean and variance  $\sigma^2$

$\varepsilon'_T$ : Independent random variables for aggregate demand in period  $T$ , normally distributed with zero mean and variance  $\sigma'^2$

$f_t$ : Forecast of disaggregate demand in period  $t$ , the forecast produced in  $t-1$  for the demand in  $t$ .

---

<sup>5</sup> Obviously other forecasting methods may also provide unbiased estimates under the stationary demand processes considered in this research but those are not considered as their analysis is beyond the scope of this research.

$F_T$  : Forecast of aggregate demand in period  $T$ , the forecast produced in  $T-1$  for the demand in  $T$ .

$\alpha$  : Smoothing constant used in Single Exponential Smoothing method before aggregation,  $0 < \alpha \leq 1$

$\beta$  : Smoothing constant used in Single Exponential Smoothing method after aggregation,  $0 < \beta \leq 1$

$MSE_{BA}$  : Theoretical Mean Squared Error ( $MSE$ ) before aggregation

$MSE_{AA}$  : Theoretical Mean Squared Error ( $MSE$ ) after aggregation

$\gamma_k$  : Covariance of lag  $k$  of disaggregate demand,  $\gamma_k = Cov(d_t, d_{t-k})$

$\gamma'_k$  : Covariance of lag  $k$  of aggregate demand,  $\gamma'_k = Cov(D_T, D_{T-k})$

$\phi$  : Autoregressive parameter before aggregation,  $|\phi| < 1$

$\phi'$  : Autoregressive parameter after aggregation,  $|\phi'| < 1$

$\theta$  : Moving average parameter before aggregation,  $|\theta| < 1$

$\theta'$  : Moving average parameter after aggregation,  $|\theta'| < 1$

$\mu$  : Expected value of disaggregate demand in any time period

$\mu'$  : Expected value of aggregate demand in any time period

It is assumed that the disaggregate demand series  $d_t$  follows a first order autoregressive moving average, ARIMA(1,0,1) or its special cases moving average order one, ARIMA (0,0,1) and an autoregressive data generation process (DGP) order one, ARIMA(1,0,0). In the following the characteristics of each process under consideration are discussed to provide the information based on the nature of the processes.

An ARIMA(1,0,1) process can be mathematically written in period  $t$  as (3-1):

$$d_t = \mu(1 - \phi) + \varepsilon_t + \phi d_{t-1} - \theta \varepsilon_{t-1}, \text{ where } |\phi| < 1, |\theta| < 1. \quad (3-1)$$

When the demand follows an ARIMA(1,0,1) process the auto-covariance and autocorrelation functions are (Box et al., 2008):

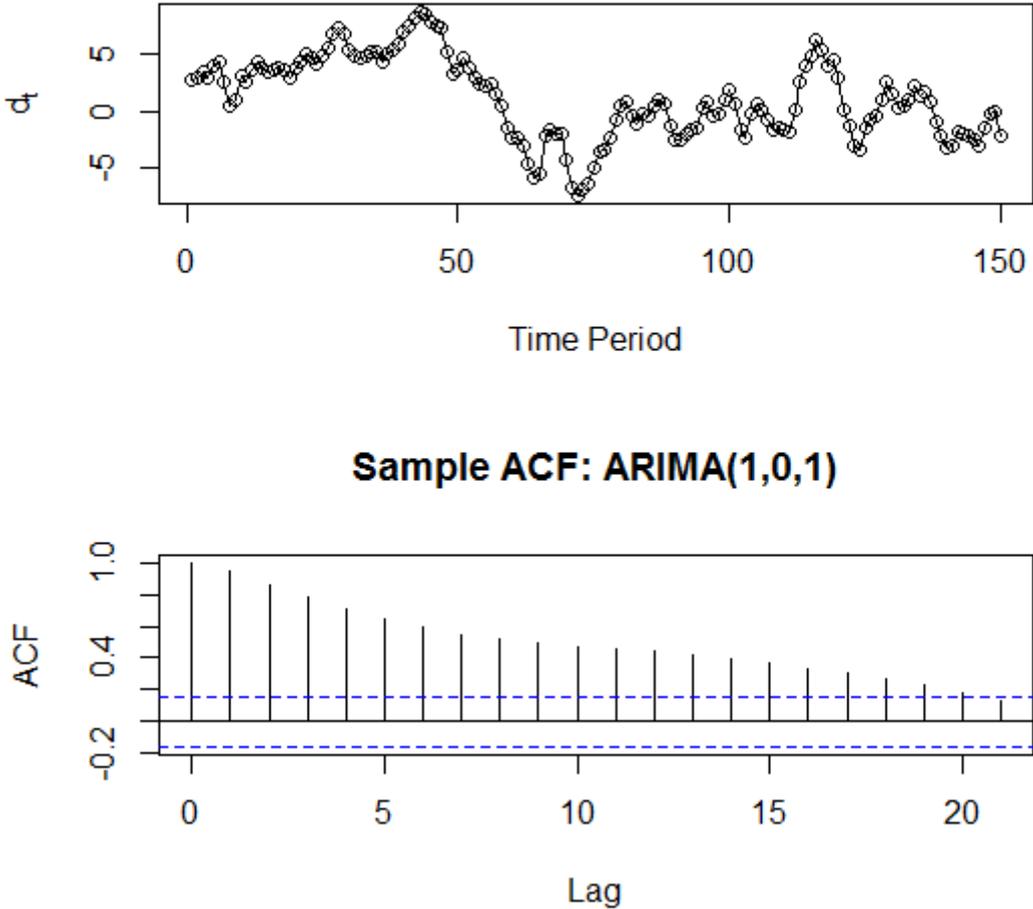
$$\gamma_k = \begin{cases} \frac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} \sigma^2 & k = 0 \\ \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \sigma^2 & |k| = 1, \\ \phi \gamma_{k-1} & |k| > 1 \end{cases} \quad (3-2)$$

$$\rho(k) = \begin{cases} \frac{(\phi - \theta)(1 - \phi\theta)}{1 - 2\phi\theta + \theta^2} & |k| = 1 \\ \phi \rho_{k-1} & |k| > 1 \end{cases} \quad (3-3)$$

For different combinations of the process parameters, the resulting underlying structure changes considerably. **Table 3-1** presents the autocorrelation structure for different process parameters which helps to better understand the process and can be useful to interpret the results of the forthcoming analysis.

**Table 3-1:** Autocorrelation of ARIMA(1,0,1) process

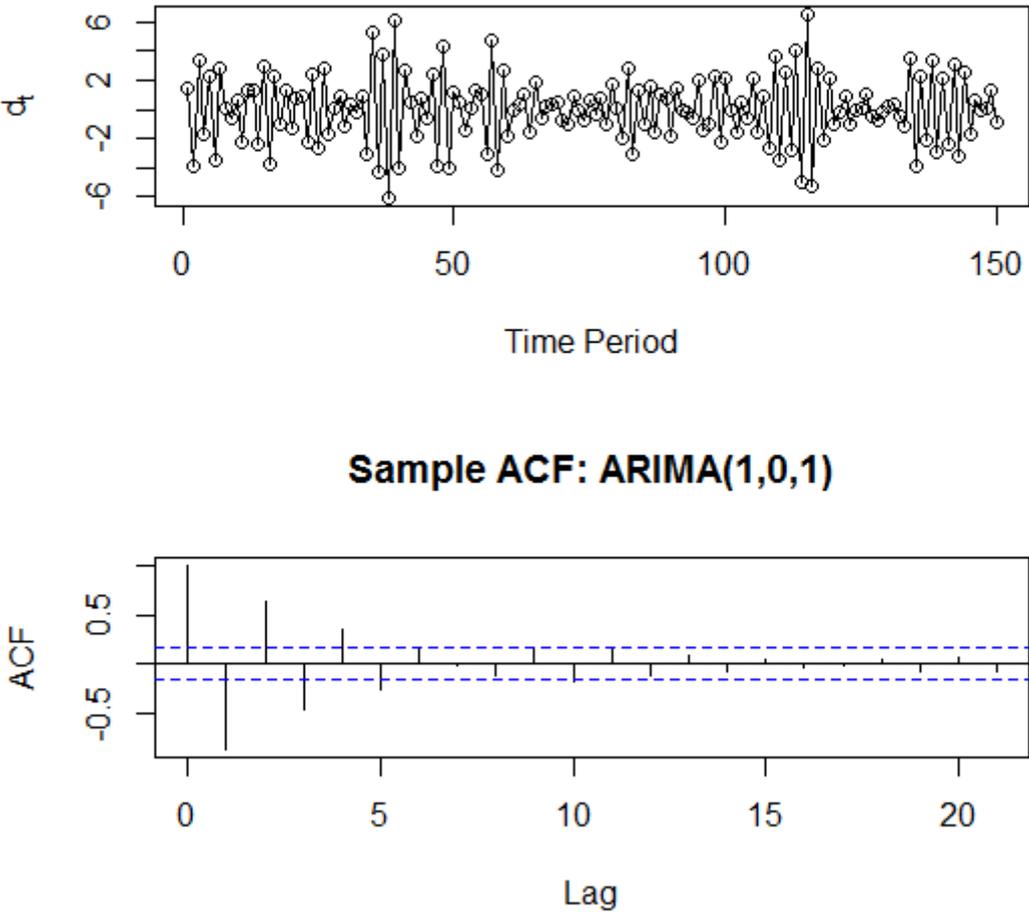
Group	Process parameter	Autocorrelation
1	$0 < \phi < 1, -1 < \theta < 0$	Always positive, $0 < \text{Autocorrelation lag} < 1$ ,
2	$-1 < \phi < 0, -1 < \theta < 0$	Oscillate between positive and negative values
3	$-1 < \phi < 0, 0 < \theta < 1$	Oscillate between positive and negative values
4	$0 < \phi < 1, 0 < \theta < 1$	For $\phi > \theta$ Always positive, $0 < \text{Autocorrelation lag} < 1$
5	$0 < \phi < 1, 0 < \theta < 1$	For $\phi < \theta$ Always negative, $-0.5 < \text{Autocorrelation lag} < 0$



**Figure 3-1:** Sample autocorrelation of ARIMA(1,0,1) process when  $\theta= -0.8$  and  $\phi=0.9$ .

**Figure 3-1** and **3-2** present the behaviour of the ARIMA(1,0,1) process for groups one and two presented in **Table 3-1**. In **Figure 3-1** it can be seen that the autocorrelation is highly positive not only for lag 1 but also for higher lags and decays exponentially. In addition it is observed that the process shape is changing slowly and there is no fluctuation between time periods.

In **Figure 3-2** the process shape is changing almost at each period and there are more fluctuations which makes the series more irregular than random series. As it can be noted that the autocorrelation decays exponentially and oscillates between positive and negative values and it tends to become zero for higher lags.



**Figure 3-2:** Sample autocorrelation of ARIMA(1,0,1) process when  $\theta= 0.8$  and  $\phi=-0.7$ .

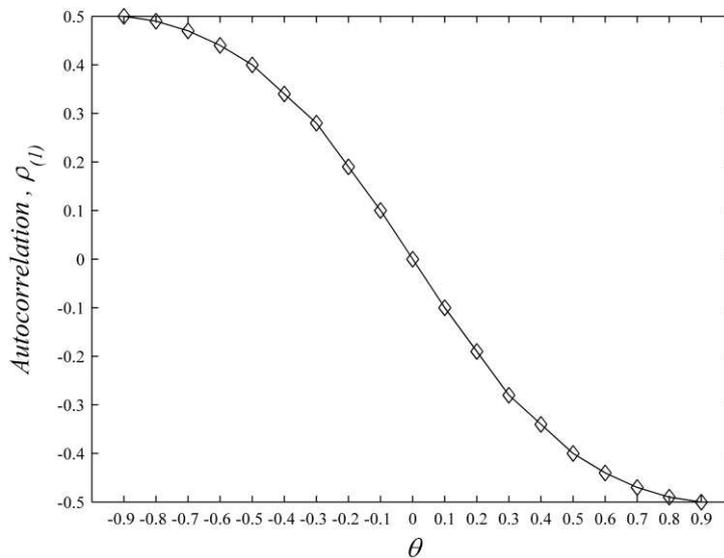
An ARIMA(0,0,1) process is a special case of a more general ARIMA(1,0,1) process where the autoregressive process is equal to zero i.e  $\phi=0$ . This process can be mathematically shown as (3-4):

$$d_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, \text{ where } |\theta| < 1, \quad (3-4)$$

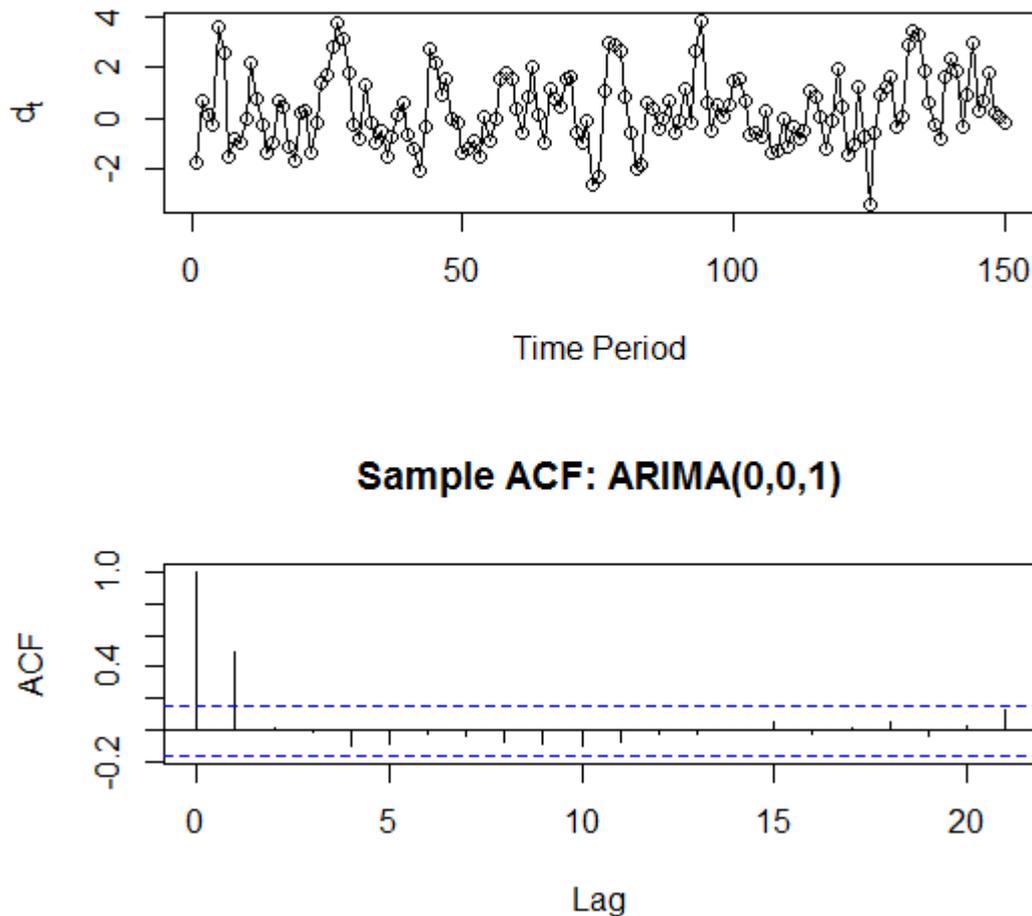
When the demand follows an ARIMA(0,0,1) process, the autocovariance and autocorrelation functions are (Wei, 2006):

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma^2 & k = 0 \\ -\theta\sigma^2 & |k| = 1 \\ 0 & |k| > 1 \end{cases} \quad (3-5)$$

$$\rho(k) = \begin{cases} \frac{-(1 + \theta^2)}{\theta} & |k| = 1 \\ 0 & |k| > 1 \end{cases} \quad (3-6)$$



**Figure 3-3:** Autocorrelation of ARIMA(0,0,1) process



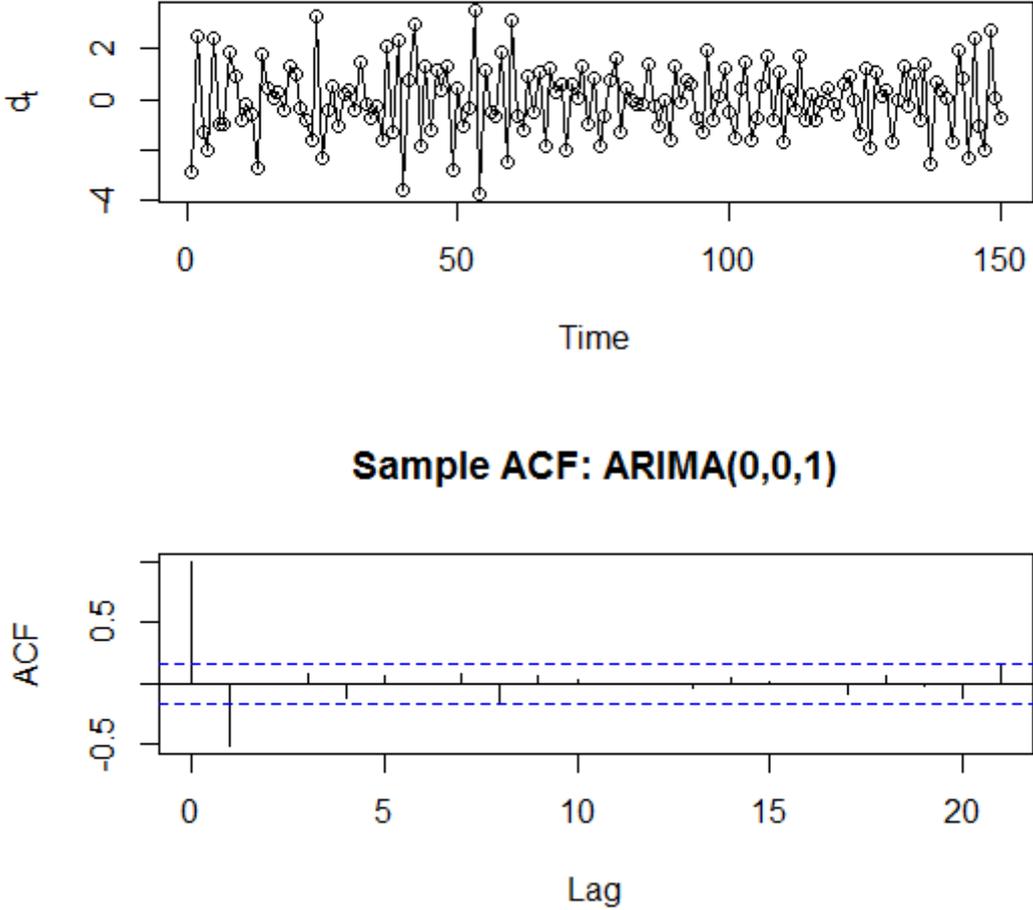
**Figure 3-4:** Sample autocorrelation and the process shape of ARIMA(0,0,1) process when  $\theta = -0.9$ .

**Figure 3-3** shows that the autocorrelations for an ARIMA(0,0,1) process varies between  $-0.5$  and  $+0.5$  for high positive and high negative values of the moving average parameter  $\theta$ , respectively. In addition the autocorrelation is equal to zero for lags greater than one.

In **Figure 3-4** the behavior of the ARIMA(0,0,1) process is presented when the moving average parameter is relatively high. It is seen that for this  $\theta$  value, the autocorrelation function is close to  $+0.5$  and the process is changing slowly. However, the rate of changing is

slower in the case of ARIMA(1,0,1) process where autocorrelation is high, this is natural as the autocorrelation function for the ARIMA(0,0,1) is much smaller than ARIMA(1,0,1) process.

When the moving average parameter takes positive values the process shape becomes more irregular compare to **Figure 3-4**. The autocorrelation function is negative for lag1 and it equals to zero for higher lags as shown in **Figure 3-5**.



**Figure 3-5:** Sample autocorrelation and the process shape of ARIMA(0,0,1) process when  $\theta=0.9$ .

Finally the Autoregressive process order one, ARIMA(1,0,0) can be represented as (3-7) which is a special case of the ARIMA(1,0,1) process where  $\theta=0$ .

$$d_t = \mu(1 - \phi) + \varepsilon_t + \phi d_{t-1}. \quad (3-7)$$

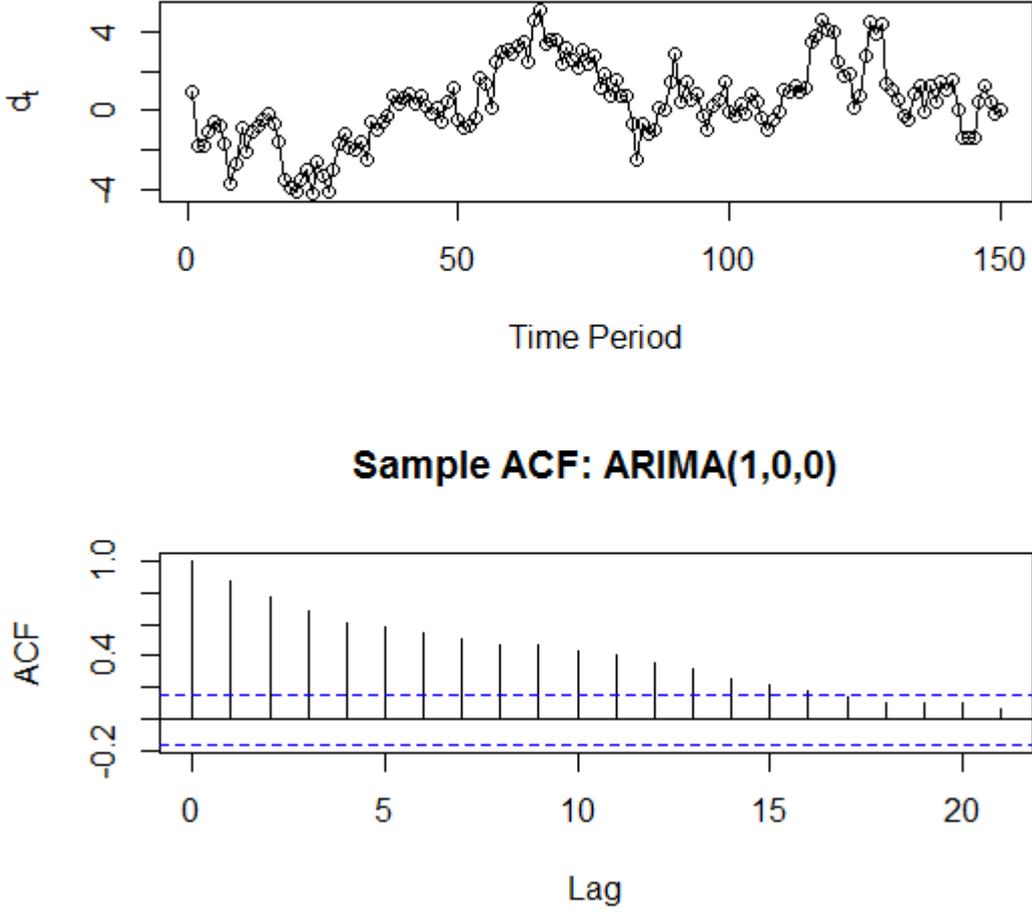
When demand follows an ARIMA(1,0,0) process the following properties exist (Wei, 2006):

$$\gamma_k = \begin{cases} \sigma^2 & k = 0 \\ 1 - \phi^2 & k = 1 \\ \phi^k \gamma_0 & k \geq 1 \end{cases}, \quad (3-8)$$

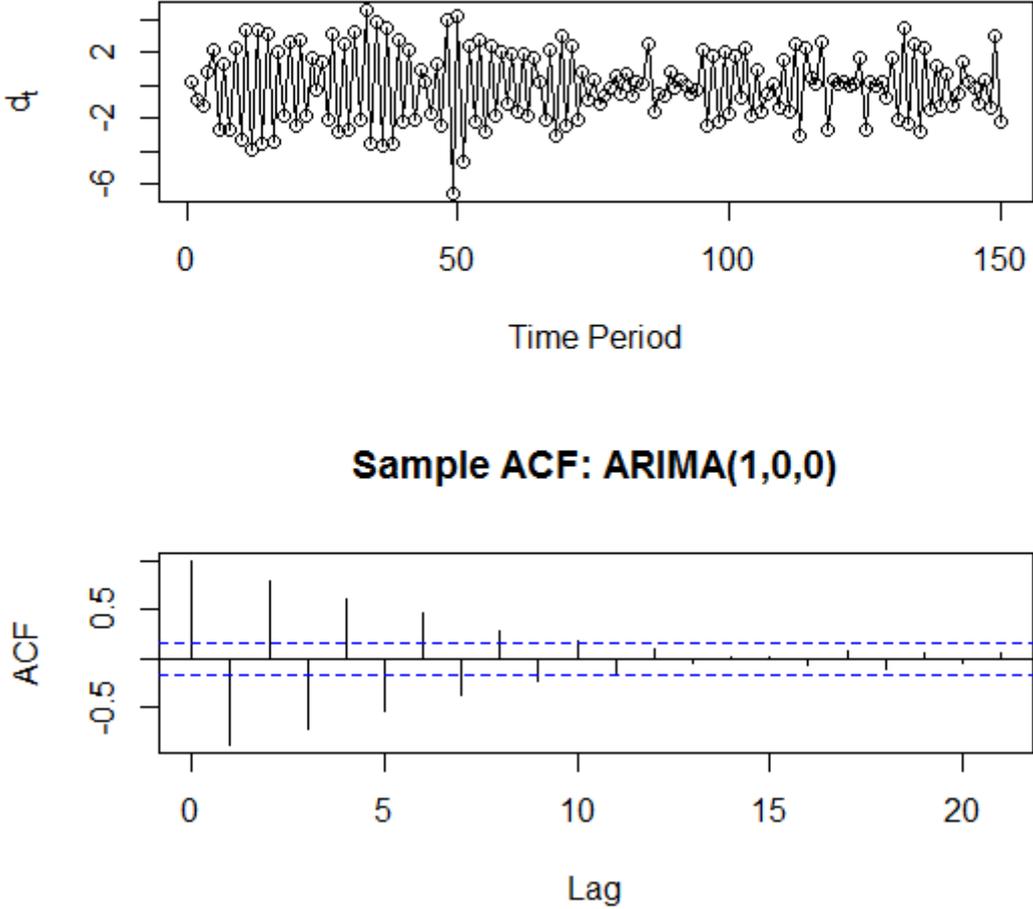
$$\rho(k) = \begin{cases} \phi^k & |k| \geq 1. \end{cases} \quad (3-9)$$

It's clear from (3-9) that when the autoregressive parameter  $\phi$  takes positive values, the autocorrelation is always positive not only for lag1 but also for higher lags periods. It exhibits a smooth exponential decay as shown in **Figure 3-6** for high positive values.

When the autoregressive parameter  $\phi$  is negative, the autocorrelation function is decays exponentially and oscillates between positive and negative values. The process shape is irregular as can be seen in **Figure 3-7**



**Figure 3-6:** Sample autocorrelation and the process shape of ARIMA(1,0,0) process when  $\phi=0.9$ .



**Figure 3-7:** Sample autocorrelation and the process shape of ARIMA(1,0,0) process when  $\phi = -0.9$ .

The  $m$  periods non-overlapping aggregate demand  $D_T$  can be expressed as a function of the disaggregate demand series as follows

$$D_{T-k} = \sum_{l=1}^m d_{t-(k-1)m-l} \quad k=1,2,\dots \quad (3-10)$$

The forecasting method considered in this study is the Single Exponential Smoothing (SES); this method is being applied in very many companies and most managers use this method in a production planning environment due to its simplicity (Gardner, 1990). Using SES, the forecast of demand in period  $t$  produced at the end of period  $t-1$  is

$$f_t = \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}, \quad (3-11)$$

It is further assumed that the standard deviation of the error term in (3-4), (3-7), and (3-10) above is significantly smaller than the expected value of the demand, so when demand is generated the probability of a negative value is negligible. Constraining  $\phi$  and  $\theta$  to lie between -1 and 1 in (3-4), (3-7), and (3-10) means that the process is stationary and invertible.

### 3.1.2 MSE derivation at disaggregate level

In this section the *MSE* of the one-step-ahead forecasts resulted from the disaggregate and aggregate demand data is derived. This section is divided into two sub-sections. First, the *MSE* before aggregation is calculated based on the direct forecast resulted from disaggregate demand. Then, the *MSE* after aggregation is configured, so the aggregate forecast is disaggregate by dividing them by aggregation level  $m$ .

#### 3.1.2.1 MSE Before Aggregation, $MSE_{BA}$

In order to calculate the  $MSE_{BA}$ , the forecasting method, SES, is directly applied to disaggregate demand data to produce one-step-ahead forecasts.

The analysis begins by deriving the  $MSE_{BA}$  for the ARIMA(1,0,1) process. As discussed above the  $MSE_{BA}$  is

$$\begin{aligned} MSE_{BA} &= \text{Var}(\text{Forecast Error}) = \\ &= \text{Var}(d_t - f_t) = \text{Var}(d_t) + \text{Var}(f_t) - 2\text{Cov}(d_t, f_t), \end{aligned} \quad (3-12)$$

Subsequently, the three parts of (3-12) should to be determined: i) variance of the demand, ii) variance of the forecast, and iii) the covariance between the demand and the forecast.

The evaluation of  $MSE_{BA}$  is begun by defining the covariance between the demand and the forecast as follows:

$$\begin{aligned} \text{Cov}(d_t, f_t) &= \text{Cov}(d_t, \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}) = \alpha \text{Cov}(d_t, \sum_{k=1}^{\infty} (1-\alpha)^{k-1} d_{t-k}) = \\ &\alpha (\text{Cov}(d_t, d_{t-1}) + (1-\alpha)\text{Cov}(d_t, d_{t-2}) + (1-\alpha)^2 \text{Cov}(d_t, d_{t-2}) + \dots), \end{aligned} \quad (3-13)$$

Considering that  $\text{Cov}(d_t, d_{t-k}) = 0$  for all  $k > 1$  and by substituting (3-2) in (3-13), the covariance between demand and its forecast is obtained:

$$\text{Cov}(d_t, f_t) = \alpha\gamma_1 + \alpha(1-\alpha)\phi\gamma_1 + \alpha(1-\alpha)^2\phi^2\gamma_1 + \dots = \frac{\alpha\gamma_1}{1-\phi+\alpha\phi}. \quad (3-14)$$

The variance of the forecast is calculated as follows:

$$\begin{aligned} \text{Var}(f_t) &= \text{Var}(\alpha d_{t-1} + (1-\alpha)f_{t-1}) = \alpha^2 \text{Var}(d_{t-1}) \\ &\quad + (1-\alpha)^2 \text{Var}(f_{t-1}) + 2\alpha(1-\alpha)\text{Cov}(d_{t-1}, f_{t-1}). \end{aligned} \quad (3-15)$$

By considering the fact that the process is stationary, it is known  $Var(f_t) = Var(f_{t-k})$  and  $Cov(d_t, f_t) = Cov(d_{t-k}, f_{t-k})$  for all  $k$  values and by substituting (3-2) and (3-14) into (3-15) the following is obtained:

$$Var(f_t) = \frac{\alpha\gamma_0}{2-\alpha} + \frac{2\alpha(1-\alpha)\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)}, \quad (3-16)$$

The equations (3-2), (3-14) and (3-16) is substituted in (3-12), these substitutions coupled with the fact that  $\gamma_0 = Var(d_{t-k})$  reveals the  $MSE_{BA}$  as follows:

$$MSE_{BA} = \frac{\sigma^2}{1-0.5\alpha} \left( \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} - \frac{\alpha(\phi-\theta)(1-\phi\theta)}{(1-\phi^2)(1-\phi+\alpha\phi)} \right). \quad (3-17)$$

As a special case, when  $\phi=0$  the ARIMA(1,0,1) becomes the ARIMA(0,0,1) process which is called MA(1) as well. Therefore, by substituting  $\phi=0$  in (3-17) the  $MSE_{BA}$  for ARIMA(0,0,1) process is obtained in (3-18):

$$MSE_{BA} = \frac{\gamma_0 - \alpha\gamma_1}{1-0.5\alpha}. \quad (3-18)$$

Autoregressive order one, ARIMA(1,0,0) or AR(1) is a special case of the ARIMA(1,0,1) process when  $\theta=0$ . Therefore, the  $MSE_{BA}$  for the ARIMA(1,0,0) process is obtained by substituting  $\theta=0$  in (3-17):

$$MSE_{BA} = \frac{\gamma_0(1-\phi)}{(1-0.5\alpha)(1-\phi+\alpha\phi)} \quad (3-19)$$

### 3.1.2.2 MSE after Aggregation, $MSE_{AA}$

In this section, the derivation of the  $MSE$  of the forecasts for the aggregation approach is determined. Disaggregate demand is first aggregate to yield high frequency demand. Then, the aggregate forecasts are provided based on the  $SES$  forecasting method. Finally, one-step-ahead estimates at the original level are given by the disaggregation of such forecasts. This disaggregation is conducted by dividing the aggregate forecast by the aggregation level  $m$ . The  $MSE_{AA}$  is defined as

$$MSE_{AA} = \text{Var}\left(d_t - \frac{F_T}{m}\right) = \text{Var}(d_t) + \frac{1}{m^2} \text{Var}(F_T) - \frac{2}{m} \text{Cov}(d_t, F_T) = \gamma_0 + \frac{1}{m^2} \text{Var}(F_T) - \frac{2}{m} \text{Cov}(d_t, F_T). \quad (3-20)$$

By applying  $SES$ , the aggregate forecast for period  $T$  is defined as

$$F_T = \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}. \quad (3-21)$$

In this section, the  $MSE_{AA}$  is derived for an  $ARIMA(1,0,1)$  demand process. When the disaggregate series follows an  $ARIMA(1,0,1)$  process, the aggregate series also follows an  $ARIMA(1,0,1)$  process but with different parameter values (Sariaslan, 2010; Tiao, 1972). The autocovariance function of an  $ARIMA(1,0,1)$  process after aggregation is given:

$$\gamma'_k = \begin{cases} \frac{1 - 2\phi'\theta' + \theta'^2}{1 - \phi'^2} \sigma'^2 & k = 0 \\ \frac{(\phi' - \theta')(1 - \phi'\theta')}{1 - \phi'^2} & |k| = 1 \\ \phi'\gamma'_{k-1} = \phi'^{k-1}\gamma_1 & |k| > 1 \end{cases} \quad (3-22)$$

From Appendix A and Based on Wei (2006) the relationship between the autocovariance function of the disaggregate and the aggregate demand for an  $ARIMA(1,0,1)$  process is obtained as follows:

$$\gamma'_k = \begin{cases} m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) & k=0 \\ \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) & |k|=1 \\ \gamma_1 \left( \phi^{m(k-1)} + 2\phi^{(k-1)m+1} + \dots + m\phi^{km-1} + (m-1)\phi^{km} + \dots + \phi^{(k+1)m-2} \right) & |k|>1 \end{cases} \quad (3-23)$$

$$\phi' = \phi^m . \quad (3-24)$$

By considering (3-23) , the autocorrelation function after aggregation is given as following:

$$\rho'(k) = \frac{\gamma'_k}{\gamma'_0} = \begin{cases} \frac{\gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right)}{m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right)} & |k|=1 \\ \phi' \rho'_{k-1} & |k|>1 \end{cases} . \quad (3-25)$$

From (B-4) and (C-4) in Appendix A and B respectively, the covariance between disaggregate demand and aggregate forecast is given in (3-26). Additionally, the variance of the aggregate forecast is given in (3-27):

$$Cov(d_t, F_T) = \frac{\beta\gamma_1}{1-\phi^m + \beta\phi^m} \times \frac{1-\phi^m}{1-\phi} , \quad (3-26)$$

$$Var(F_T) = \frac{\beta\gamma'_0}{2-\beta} + \frac{2\beta(1-\beta)\gamma'_1}{(2-\beta)(1-\phi' + \beta\phi')} . \quad (3-27)$$

Now, the equations (3-26) and (3-27) are substituted in (3-20). Then, the equations (3-23) and (3-2) are substituted in that result. Finally, the  $MSE$  of the forecast after aggregation is given as follows:

$$\begin{aligned}
 MSE_{AA} = & \left( \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} \sigma^2 \right) + \frac{1}{m^2} \left( \beta \left( \frac{m \left( \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} \sigma^2 \right) + \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \sigma^2 \right) \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right)}{2-\beta} \right) \right. \\
 & \left. + \frac{2\beta(1-\beta) \left( \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \sigma^2 \right) \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} \right) \\
 & - \frac{2}{m} \left( \frac{\beta\gamma_1}{1-\phi^m + \beta\phi^m} \times \frac{1-\phi^m}{1-\phi} \right).
 \end{aligned} \tag{3-28}$$

As a special case, when  $\phi=0$  the ARIMA(1,0,1) process becomes an ARIMA(0,0,1) process which is also called MA(1), therefore the  $MSE_{AA}$  for the ARIMA(0,0,1) process is obtained by substituting  $\phi=0$  in (3-28) :

$$MSE_{AA} = \left( \gamma_0 + \frac{1}{m^2} \left( \frac{\beta m \gamma_0 + 2m\beta\gamma_1 - 2\beta^2\gamma_1}{2-\beta} \right) - \frac{2\beta\gamma_1}{m} \right). \tag{3-29}$$

To obtain the  $MSE_{BA}$  for the ARIMA(1,0,0) or AR(1) process,  $\theta=0$  is substituted in (3-28), therefore, the  $MSE_{BA}$  for the ARIMA(1,0,0) process is:

$$MSE_{AA} = \gamma_0 \left( \begin{array}{l} 1 + \frac{1}{m^2} \left( \frac{\beta \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right)}{2-\beta} + \frac{2\beta(1-\beta) \left( \sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} \right) \\ - \frac{2}{m} \left( \frac{\beta\phi}{1-\phi^m + \beta\phi^m} \times \frac{1-\phi^m}{1-\phi} \right) \end{array} \right) \quad (3-30)$$

### 3.1.3 MSE derivation at aggregate level

In this section, the variance of the error of the cumulative  $m$ -step-ahead forecast is derived. Firstly, the  $MSE$  of the forecasts resulted from the disaggregate demand data,  $MSE_{BA}$ , is calculated. Then, the aggregate demand is used to calculate the aggregate forecasts and, consequently, the  $MSE_{AA}$  is obtained.

#### 3.1.3.1 MSE Before Aggregation, $MSE_{BA}$

The analysis begins by deriving the  $MSE_{BA}$  for the ARIMA(1,0,1) process. The  $MSE$  of the forecasts for the non-aggregation approach is derived as follows: Firstly, one step ahead demand forecasts are obtained based on the  $SES$  method. Then, the results are multiplied by the aggregation level  $m$ . This results in cumulative  $m$ -step-ahead estimates at the aggregate level. The  $MSE_{BA}$  is defined by:

$$MSE_{BA} = Var(D_T - mf_t) = Var(D_T) + m^2 Var(f_t) - 2m Cov(D_T, f_t), \quad (3-31)$$

In this section, the  $MSE_{BA}$  is derived for an ARIMA(1,0,1) demand process.

As it is defined in (3-16), the variance of the disaggregate forecast is:

$$Var(f_t) = \frac{\alpha\gamma_0}{2-\alpha} + \frac{2\alpha(1-\alpha)\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)}, \quad (3-32)$$

From (D-4) in Appendix D, the covariance between aggregate demand and disaggregate forecast is given as follows:

$$Cov(D_T, f_t) = \frac{\alpha\gamma_1[(1-\phi) + \phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)}. \quad (3-33)$$

When the disaggregate series follows an ARIMA(1,0,1) process, the aggregate series also follows an ARIMA(1,0,1) process but with different parameter values (Brewer, 1973; Sariaslan, 2010). The aggregate demand is represented as follows:

$$D_T = \mu'(1-\phi') + \varepsilon'_T + \phi' D'_{T-1} - \theta'\varepsilon'_{T-1}, \text{ where } |\theta'| < 1, |\phi'| < 1. \quad (3-34)$$

The relationship between the disaggregate and the aggregate process parameters is given in (3-23).

By considering that  $\gamma'_0 = Var(D_T)$  and substituting (3-16), (3-33) and (3-23) into (3-31) the following equation is given:

$$MSE_{BA} = m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) + m^2 \left( \frac{\alpha\gamma_0}{2-\alpha} + \frac{2\alpha(1-\alpha)\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)} \right) - \frac{2m\alpha\gamma_1[(1-\phi) + \phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)} \quad (3-35)$$

Finally, by substituting (3-2) into (3-35), the  $MSE_{BA}$  is obtained:

$$MSE_{BA} = \sigma^2 \left( \begin{aligned} & \left( \frac{m(1-2\phi\theta + \theta^2)}{1-\phi^2} \right) + \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \\ & + \frac{m^2\alpha}{2-\alpha} \left( \frac{1-2\phi\theta + \theta^2}{1-\phi^2} \right) + \frac{2m^2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \\ & - \frac{2m\alpha[(1-\phi)+\phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \end{aligned} \right) \quad (3-36)$$

As a special case, by substituting  $\phi=0$  in (3-36) the  $MSE_{BA}$  for the ARIMA(0,0,1) process is given as follows:

$$MSE_{BA} = m(1+\theta^2)\sigma^2 - 2(m-1)\theta\sigma^2 + m^2 \left( \frac{\alpha(1+\theta^2)\sigma^2 - 2\alpha(1-\alpha)\theta\sigma^2}{(2-\alpha)} \right) + 2m\alpha\theta\sigma^2. \quad (3-37)$$

The  $MSE_{BA}$  for the ARIMA(1,0,0) process is obtained by substituting  $\theta=0$  in (3-36):

$$MSE_{BA} = \frac{\sigma^2}{1-\phi^2} \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right) + m^2 \left( \frac{\alpha(1+\phi-\alpha\phi)\sigma^2}{(2-\alpha)(1-\phi+\alpha\phi)(1-\phi^2)} \right) - \frac{2m\alpha\phi\sigma^2((1-\phi)+\phi(1-\phi^{m-1}))}{(1-\phi)(1-\phi+\alpha\phi)(1-\phi^2)}. \quad (3-38)$$

### 3.1.3.2 MSE after Aggregation, $MSE_{AA}$

In this section, the  $MSE$  of the cumulative  $m$  step ahead forecast is obtained from the aggregate demand data.

In this section, The  $MSE_{AA}$  is calculated for the ARIMA(1,0,1) process. The  $MSE_{AA}$  is defined as:

$$MSE_{AA} = Var(D_T - F_T) = Var(D_T) + Var(F_T) - Cov(D_T, F_T), \quad (3-39)$$

From (C-3) and (C-4) in Appendix C, respectively, the covariance between the aggregate demand and its forecast is given as following:

$$\text{Cov}(D_T, F_T) = \frac{\beta\gamma'_1}{1 - \phi' + \beta\phi'} \quad (3-40)$$

$$\text{Var}(F_T) = \frac{\beta\gamma'_0}{2 - \beta} + \frac{2\beta(1 - \beta)\gamma'_1}{(2 - \beta)(1 - \phi' + \beta\phi')} \quad (3-41)$$

Now by considering that  $\gamma'_0 = \text{Var}(D_T)$  and substituting (3-40) and (3-41) in (3-39), the  $MSE_{AA}$  is obtained as follows:

$$MSE_{AA} = \frac{2\gamma'_0}{2 - \beta} - \frac{2\beta\gamma'_1}{(2 - \beta)(1 - \phi' + \beta\phi')} \quad (3-42)$$

By substituting (3-23) and (3-24) into (3-42), the following equation is given:

$$MSE_{AA} = \frac{2 \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right)}{2 - \beta} - \frac{2\beta \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right)}{(2 - \beta)(1 - \phi^m + \beta\phi^m)} \quad (3-43)$$

Finally, by substituting (3-2) into (3-43), the  $MSE_{AA}$  becomes:

$$MSE_{AA} = \frac{2 \left( \frac{m(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} + \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right) \right)}{2 - \beta} - \frac{2\beta \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right)}{(2 - \beta)(1 - \phi^m + \beta\phi^m)} \quad (3-44)$$

The  $MSE_{AA}$  for the ARIMA(0,0,1) process is obtained by substituting  $\phi=0$  in (3-44):

$$MSE_{AA} = \frac{\beta(m(1+\theta^2)\sigma^2 - 2(m-1)\theta\sigma^2) - 2\beta(1-\beta)\theta\sigma^2}{2-\beta} + 2\beta\theta\sigma^2 + m(1+\theta^2)\sigma^2 - 2(m-1)\theta\sigma^2. \quad (3-45)$$

The  $MSE_{AA}$  for the ARIMA(1,0,0) process is obtained by substituting  $\phi=0$  in (3-44):

$$MSE_{AA} = \frac{\sigma^2}{1-\phi^2} \left( \begin{aligned} & \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right) + \frac{\beta \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right)}{2-\beta} \\ & + \frac{2\beta(1-\beta) \left( \sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} - \frac{2\beta \left( \sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right)}{1-\phi^m + \beta\phi^m} \end{aligned} \right). \quad (3-46)$$

## 2. Comparative analysis

The effectiveness of temporal aggregation as compared to non-aggregation may be assessed by analyzing the ratio of their variance of the forecast error or, equivalently, their  $MSE$ s. Recall from section 1, that a value of  $MSE_{BA}/MSE_{AA}$  greater than one implies that the aggregation approach is superior to the non-aggregation one, whereas a value that is lower than one implies the opposite. A ratio value equal to one means that performance is the same.

In section 3.2.1, the impact of the aggregation level,  $m$ , the smoothing constant values,  $\alpha$  and  $\beta$ , the moving average parameter,  $\theta$ , and the autoregressive parameter,  $\phi$ , on the ratio of  $MSE_{BA}/MSE_{AA}$  is investigated by varying their values. In section 3.2.2, the conditions under which one approach outperforms the other are analytically determined. Finally in Sub-section 3.2.3 the determination of the optimum aggregation level is considered.

### 3.2.1 Impact of the parameters – sensitivity analysis

In this Sub-section the effect of the parameters  $m, \alpha, \beta, \theta$ , and  $\phi$  on the ratio  $MSE_{BA}/MSE_{AA}$  is analysed. Note that  $m, \alpha, \beta$ , are control parameters often set by the

forecaster, whereas  $\theta$  and  $\phi$  are parameters associated with the underlying demand generation process (process parameters). Therefore, it is interesting to know which values of the control parameters lead to a ratio higher than one, for any given values of the process parameters. In real world settings, data could typically be aggregate as weekly ( $m=7$ ) from daily data, yearly ( $m=4$ ) from quarterly, monthly ( $m=4$ ) from weekly, quarterly ( $m=3$ ) from monthly, semi-annually ( $m=6$ ) from monthly and annually ( $m=12$ ) from monthly data or it may also be aggregate at some other level to reflect relevant business concerns (e.g. equal to the lead time length). Given the considerable number of control parameter combinations, it is natural that only some results may be presented here.

### 3.2.1.1 Comparison at disaggregate level

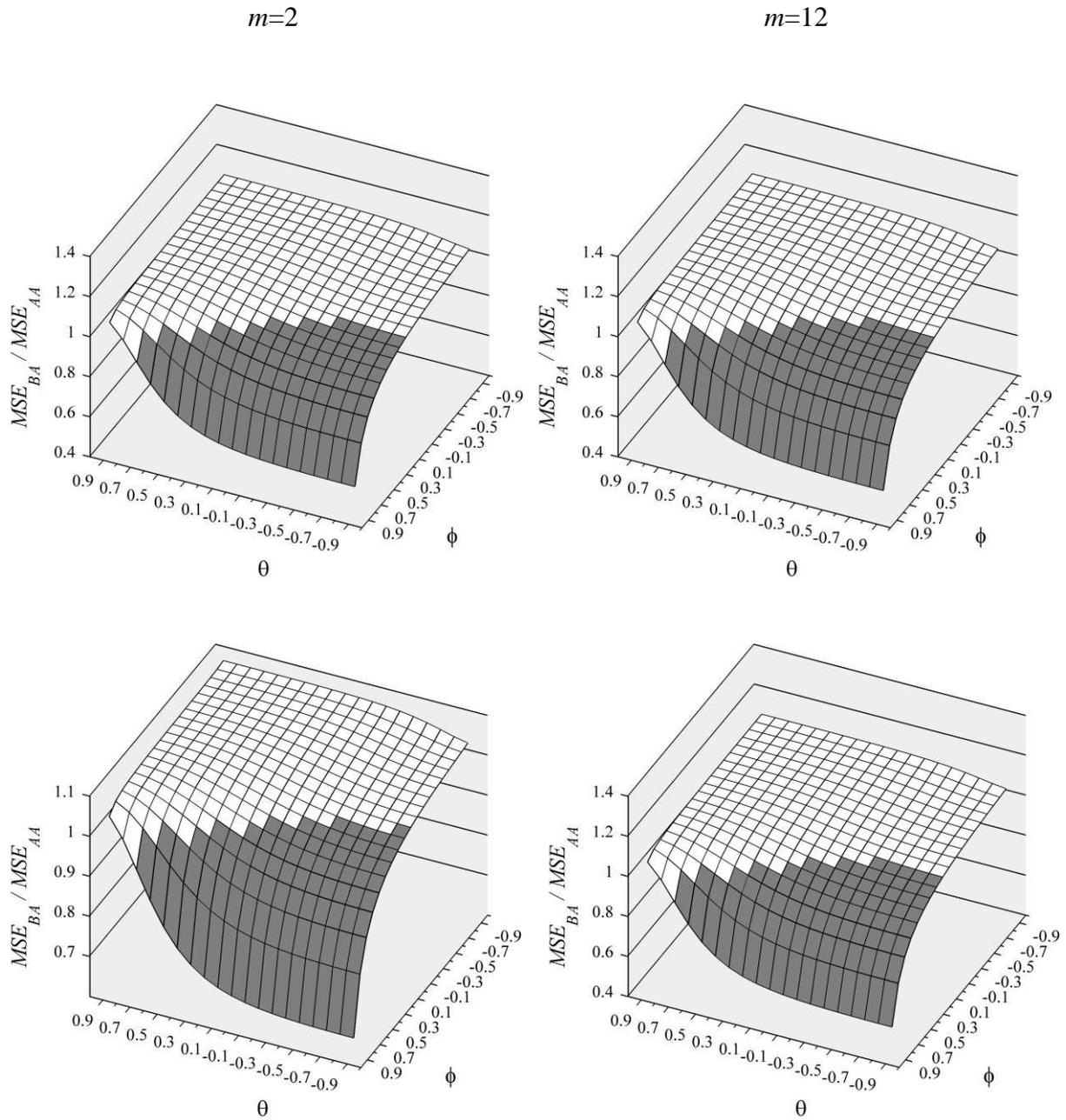
In this sub-section, the impact of the parameters on the ratio is evaluated when comparing at the disaggregate level.

#### 3.2.1.1.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

In this sub-section the effect of the parameters  $m, \alpha, \beta, \theta$  and  $\phi$  on the ratio of  $MSE_{BA}/MSE_{AA}$  is evaluated when the non-aggregate demand follows an ARIMA(1,0,1) process. We attempt to intuitively explain the effect of these parameters on the ratio. The aggregation level,  $m$  between 2 and 24,  $-0.9 \leq \theta \leq 0.9$  with increments of 0.1,  $-0.9 \leq \phi \leq 0.9$  with increments of 0.1,  $0 < \alpha < 1$  with increments of 0.05 and  $0 < \beta < 1$  with increments of 0.05 is considered.

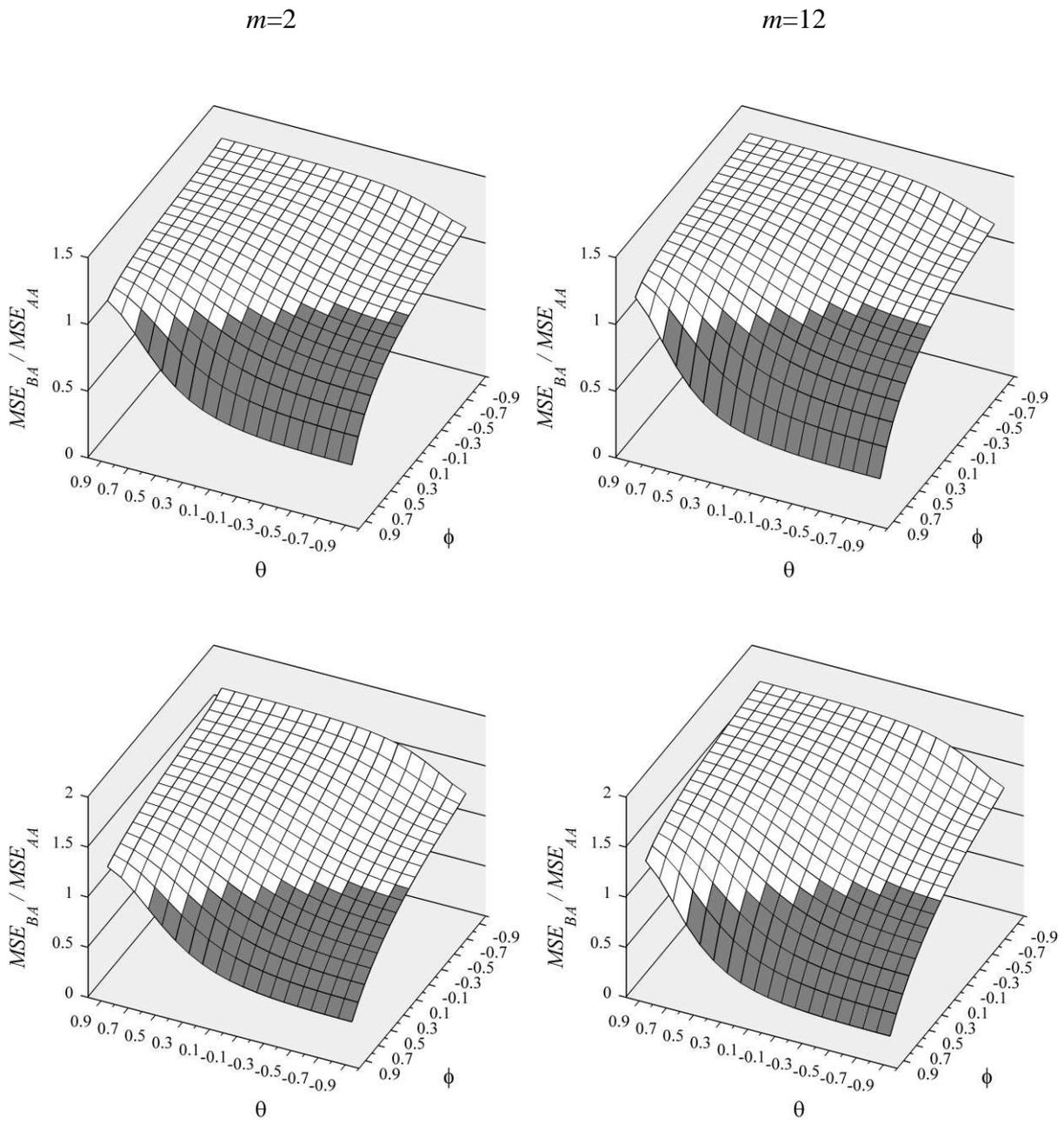
Figure 3-8 presents the impact of the parameters on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$ ,  $\alpha = 0.1$ , and  $\beta = 0.01, 0.05$ , moreover Figure 3-9 shows this impact for  $m = 2, 12$ ,  $\alpha = 0.3, \beta = 0.1$ , and  $\alpha = 0.5, \beta = 0.2$  when the non-aggregate demand series follows an ARIMA(1,0,1) process. Shaded areas represent a behavior in favor of the non-aggregation approach. It is seen that the superiority of each approach is a function of  $m, \phi, \theta, \alpha$  and  $\beta$ . The analysis shows that for a fixed value of the smoothing constants, increasing the aggregation level improves the accuracy of the aggregation approach. Additionally, for a

fixed aggregation level  $m$  and smoothing constant before aggregation  $\alpha$ , increasing  $\beta$  values decreases the performance of the aggregation approach.



**Figure 3-8:** Impact of  $m$ ,  $\theta$ ,  $\phi$ ,  $\alpha$  and  $\beta$  on the ratio of MSE:  $\alpha = 0.1, \beta = 0.01$  (top)

$\alpha = 0.1, \beta = 0.05$  (Bottom)



**Figure 3-9:** Impact of  $m, \theta, \phi, \alpha$  and  $\beta$  on the ratio of MSE:  $\alpha = 0.3, \beta = 0.1$ (top)

$\alpha = 0.5, \beta = 0.2$ (bottom)

In both Figure 3-8 and (3-9), it is revealed that for high positive values of the moving average parameter  $\theta$  and for high negative values of the autoregressive parameters  $\phi$ , the aggregation approach always yields more accurate forecasts than the non-aggregation approach. However, when  $\theta$  takes negative values and  $\phi$  takes positive values, the non-aggregation approach outperforms the aggregation one. By referring to **Table 3-1** it is obvious that the latter case corresponds to the high positive autocorrelation. Meanwhile, in the former one, the autocorrelation is not always positive and it oscillates between positive and negative values. Therefore, for a high positive autocorrelation the aggregation approach does not work and the non-aggregation approach provides more accurate results. This is generally true despite the varying of the control parameters. Thus, the aggregation approach is not recommended when autocorrelation is highly positive and associated with smaller values of  $\beta$  generally smaller or equal to  $\alpha$ . The analysis shows that even for high values of the aggregation level the area in which aggregation does not work remains almost unchanged.

Generally, as  $\theta$  gets more negative and  $\phi$  gets positive in the ARIMA(1,0,1) process, the correlation between two consecutive demand  $d_t$  gets larger. Note that for the ARIMA(1,0,1) process the autocorrelation spans all time lags (not only lag 1). Therefore, for highly positive values of  $\phi$  and highly negative values of  $\theta$ , the correlation between the consecutive and non-consecutive periods becomes extremely positive. As a result, when the demand series are highly positive correlated no level of aggregation can improve the accuracy of forecasts.

In Appendix E, it is revealed that the non-overlapping temporal aggregation approach reduces the demand variability of the ARIMA(1,0,1) process. Additionally, by increasing aggregation level more reduction in coefficient of variation can be obtained.

It can be shown that applying non-overlapping temporal aggregation decreases the value of the autocorrelation function. Moreover, increasing the aggregation level leads to more reduction in autocorrelation and it becomes close to zero for high values of  $m$ . As a result, the aggregate series becomes similar to a white noise process and it is almost random. Therefore, for the aggregate series, the smaller value of the smoothing constant  $\beta$ , generally smaller than  $\alpha$ , should be selected.

When the disaggregate process follows an ARIMA(1,0,1), the aggregation approach can reach the more accurate results when the aggregation level,  $m$ , is high and the smoothing constant after aggregation  $\beta$  is low and smaller than  $\alpha$ , generally for high positive autocorrelation, aggregation approach is not recommended.

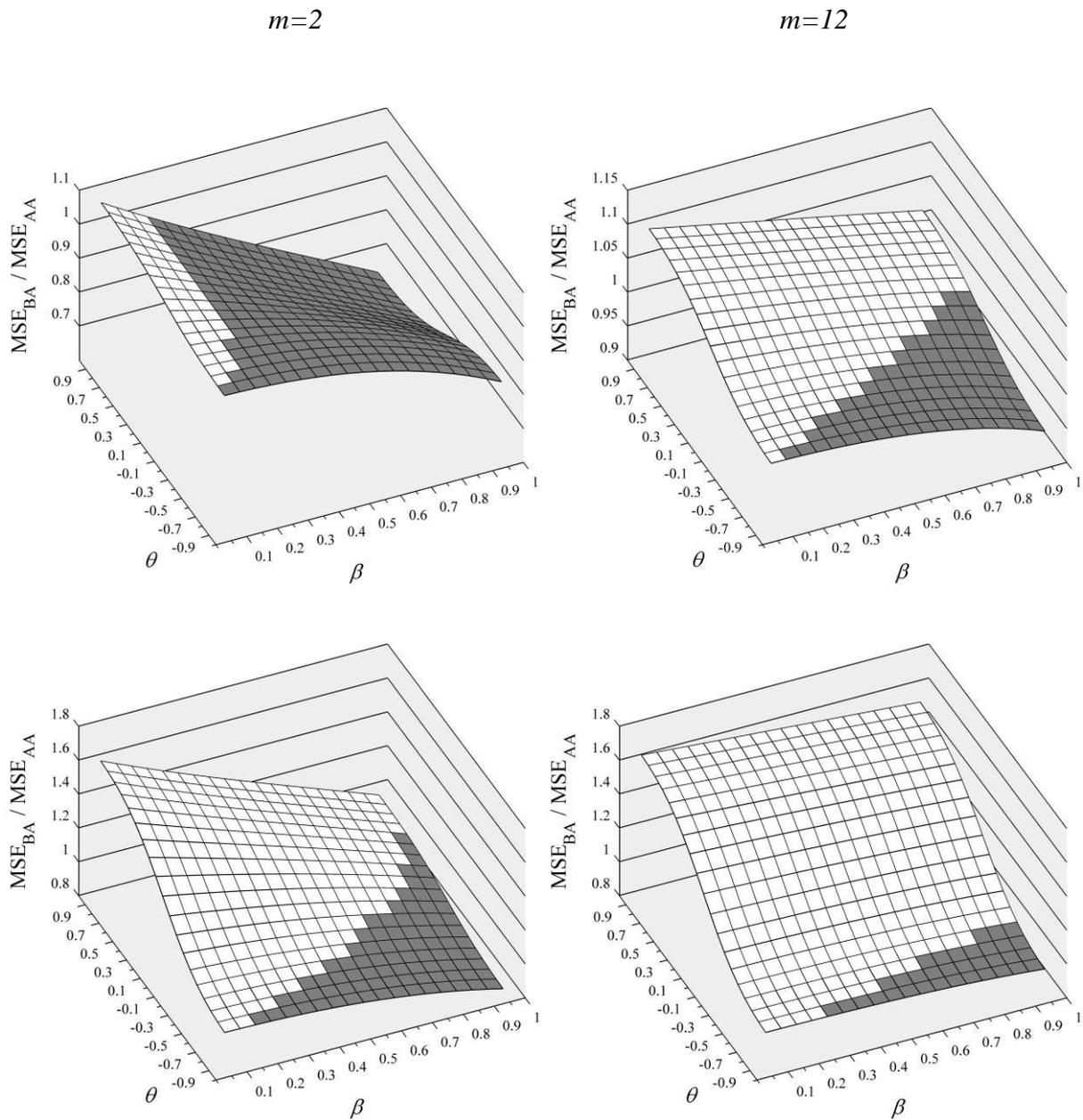
The presented results in this section show that the selection of control parameters influences the superiority of each approach and this superiority is a function of all parameters  $m$ ,  $\alpha$ ,  $\beta$ ,  $\theta$  and  $\phi$ , therefore in the section 3.2.2 we determine theoretically the conditions under which each approach outperforms another one when the disaggregate demand series follows an ARIMA(1,0,1) process.

### 3.2.1.1.2 Moving average process order one, ARIMA(0,0,1)

Figure 3-10 presents the impact of the control parameter  $\beta$  on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$  and  $\alpha = 0.1, 0.5$ , when the disaggregate demand series follows an ARIMA(0,0,1) process. Shaded areas represent a behavior in favor of the non-aggregation approach.

The results show that for a fixed value of  $\alpha$ , by increasing the aggregation level, the aggregation approach provides more accurate forecasts than the non-aggregation one. On the other hand, when considering a fixed value of the aggregation level, increasing  $\beta$  results in a deterioration of the aggregation approach. If the selected smoothing constant value after aggregation,  $\beta$ , is considerably higher than the smoothing constant used in the original data,  $\alpha$ , then the aggregation approach is not preferable. Alternatively, the aggregation approach may produce more accurate forecasts unless  $\theta$  takes highly negative values.

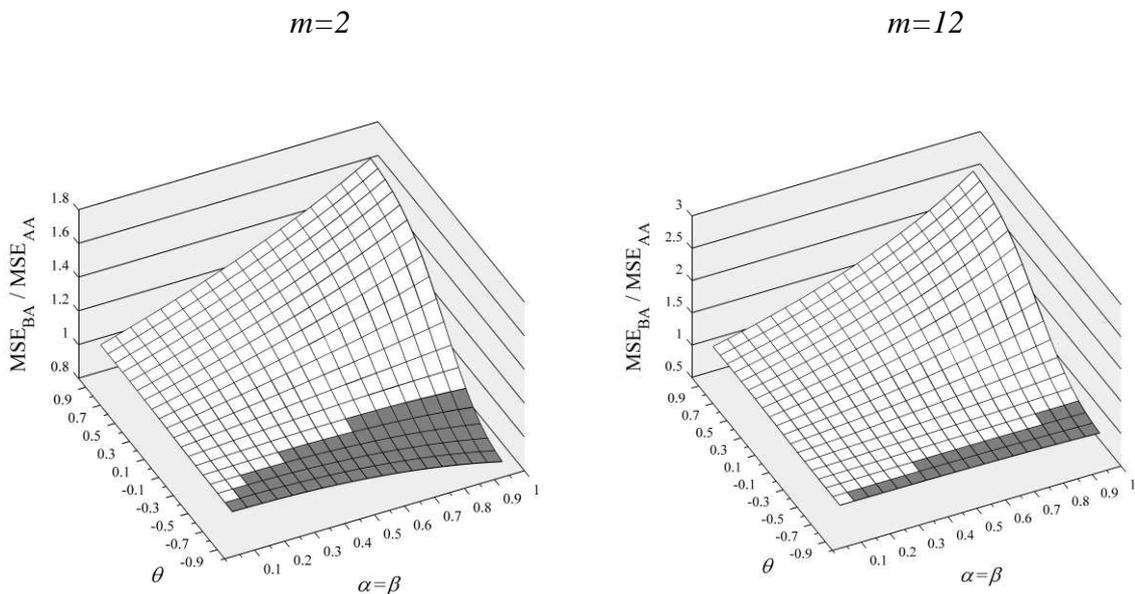
In the particular case where the smoothing constant parameters before and after aggregation are identical ( $\alpha = \beta$ ), the aggregation approach outperforms the non-aggregation one in all cases, except those associated with high negative values of  $\theta$  (high positive autocorrelation). Moreover, even in those cases, when increasing the aggregation level the performance of the aggregation approach is improved.



**Figure 3-10:** Impact of  $m$ ,  $\theta$ ,  $\alpha$  and  $\beta$  on the MSE ratio for  $\alpha = 0.1$  (top) and  $\alpha = 0.5$  (bottom)

The impact of the smoothing parameter  $\beta$  and the aggregation level  $m$  is quite intuitive similar to the ARIMA(1,0,1) process. In fact, it is obvious that the coefficient of variation (CV) of the non-overlapping temporally aggregate demand is smaller than the CV of the original (disaggregate demand) and it can be shown that by increasing the aggregation level the coefficient of variation of demand is further reduced. This means that high aggregate order

series are associated with less dispersion than low aggregate order series. In addition, by considering the autocovariance function before and after aggregation for the ARIMA(0,0,1) process, it is seen that the application of the non-overlapping temporal aggregation decreases the value of the autocorrelation function. Additionally, increasing the aggregation level leads to a higher reduction in the autocorrelation which eventually becomes zero for high aggregation level. That is, the aggregate series has a tendency towards a white noise process in which case small values of the smoothing constant lead to smaller *MSEs*. Therefore, setting  $\beta$  to be small ( $\beta$  should be smaller than  $\alpha$ ) in conjunction with high aggregation levels provides an advantage to the aggregation approach. This is confirmed by the results presented in **Figure 3-10**.



**Figure 3-11:** Impact of control parameters for ARIMA(0,0,1) process on the MSE ratio

It should be noted that even if the selected  $\beta$  is smaller than  $\alpha$ , there are some cases in which the aggregation approach is not preferable. This can be attributed to the potential high positive autocorrelation between demand periods. For negative values of  $\theta$ , the autocorrelation is positive; for positive values of  $\theta$  the autocorrelation is negative and for the white noise process, the autocorrelation is zero. An aggregation of highly positive

autocorrelated series does not add as much value as aggregating series with less positive autocorrelation. However, for very high aggregation level, the aggregation approach may outperform the non-aggregation one even for high positive autocorrelation.

When the non-aggregate items follow an ARIMA(0,0,1) process, as  $\theta$  gets more negative, the correlation between two consecutive  $d_t$  increases. For an ARIMA(0,0,1) process the only autocorrelation is autocorrelation lag 1 and for other lags, it is equal to zero. The value of autocorrelation lag 1 is varying between -0.5 and 0.5.

$$\rho = \begin{cases} -0.5 < \rho < 0.5 & |k| = 1 \\ 0 & |k| > 1 \end{cases} \quad (3-47)$$

It can be observed that the maximum positive autocorrelation of lag1 for an ARIMA(0,0,1) process is around 0.5 while this value almost equals to one for an ARIMA(1,0,1) process. These examples show that the performance superiority of each approach is a function of all the control and the process parameters. The selection of the control parameters  $\alpha$ ,  $\beta$  and  $m$ , influence the effectiveness of the aggregation approach in conjunction with the consideration of the process parameters. In sub-section 3.2.2.1.2 the conditions under which each approach produces more accurate forecasts for a fixed value of  $\alpha$  are identified.

### 3.2.1.1.3 Autoregressive process order one, ARIMA(1,0,0)

When the non-aggregate demand series follows an ARIMA(1,0,0) process, the impact of the control parameters  $m$ ,  $\alpha$ ,  $\beta$  on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$  and  $\alpha = 0.1, 0.5$  is presented as can be seen in Figure 3-12. Similar to the cases of the ARIMA(1,0,1) and the ARIMA(0,0,1) processes, it can be seen that the superiority of each approach is a function of all the control and the process parameters. The results show that for a fixed value of  $\alpha$ , increasing the aggregation level results in an improvement in the accuracy of the aggregation approach. Conversely, for a fixed aggregation level, increasing  $\beta$  results

in a deterioration of the performance. In addition,  $\beta$  should be generally smaller than  $\alpha$  in order for the aggregation approach to produce more accurate forecasts.

Figure 3-12 shows that for highly positive values of the autoregressive parameter  $\phi$  the aggregation approach does not work well and the non-aggregation approach provides more accurate results. This is generally true regardless of the values employed by the other control parameters. Therefore, the aggregation approach is not recommended in such cases.

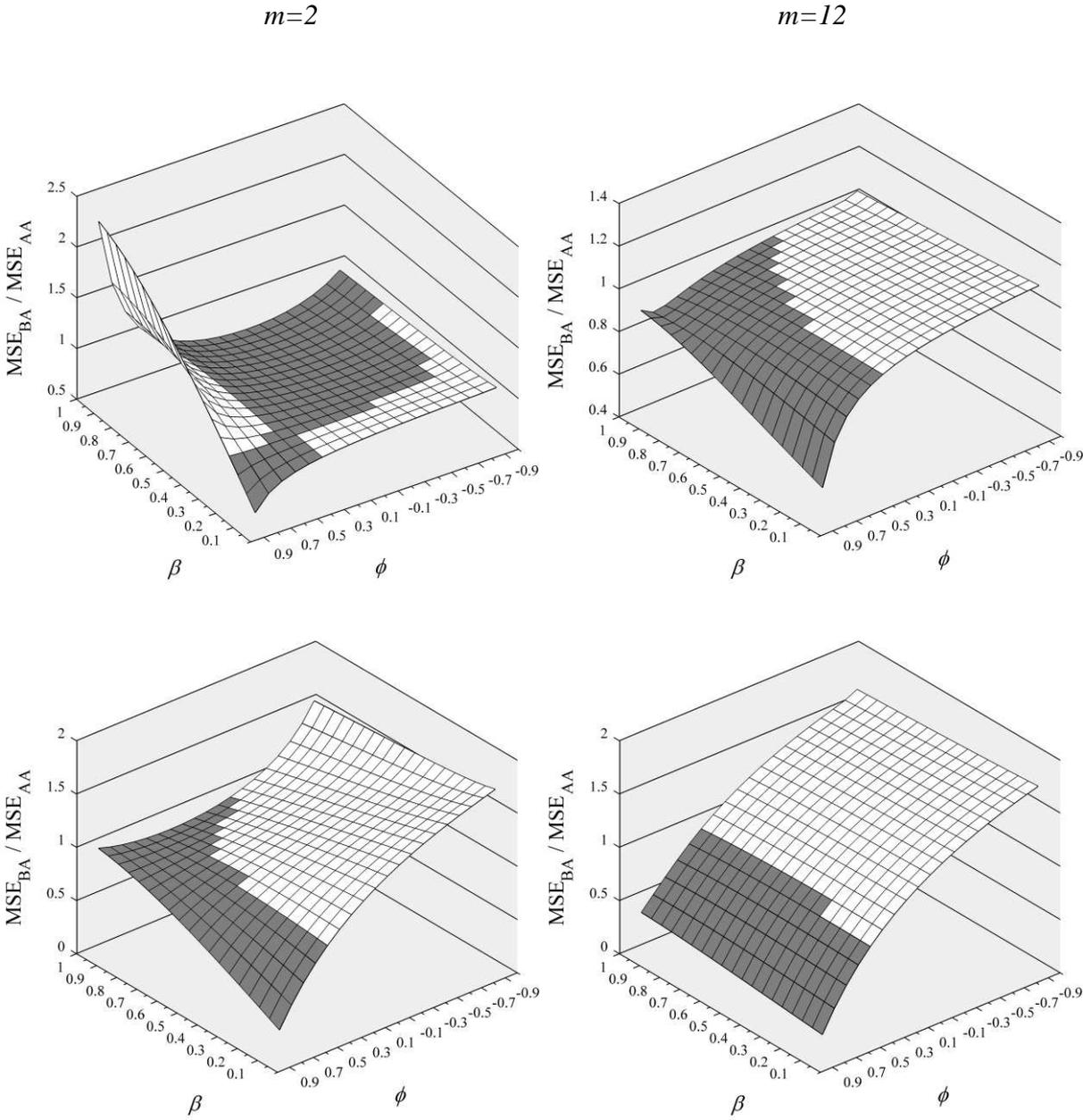
When the smoothing constant parameters before and after aggregation are identical (i.e.  $\alpha = \beta$ ), the aggregation approach outperforms the non-aggregation one in all cases, except those associated with highly positive values of  $\phi$ . In those exceptional cases the comparative performance of the two approaches is insensitive to the increase of the aggregation level and even for very high aggregation levels, no improvement is observed for the aggregation approach. The impact of the smoothing parameter  $\beta$  and the aggregation level  $m$  on the ratio is similar to that reported for the ARIMA(0,0,1) and ARIMA(1,0,1) processes.

When  $\phi$  is positive for an ARIMA(1,0,0) process, the series is 'slowly changing' or can be considered as a positively autocorrelated process. In addition, when the non-aggregate demand follows an ARIMA(1,0,0) process, the autocorrelation spans all time lags (not only lag 1). Therefore, for highly positive values of  $\phi$ , the correlation between the consecutive and non-consecutive periods becomes very high as can be obtained in (3-48) .

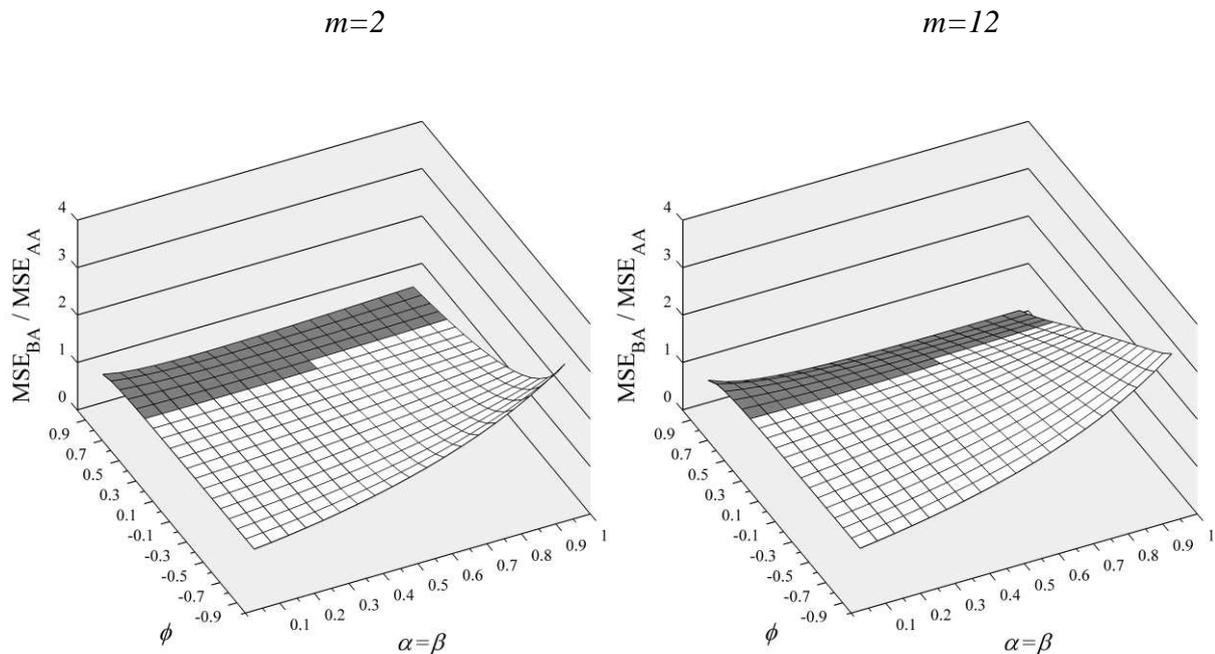
$$\rho(k) = \phi^k, \quad \text{for all } k. \quad (3-48)$$

For instance, for lag1, the autocorrelation values vary between -1 and +1,  $-1 < \rho(1) < +1$ .

It can be seen that the maximum positive autocorrelation of lag1 is around +1 for ARIMA(1,0,0) process. Consider a case where the autocorrelation = 1, say,  $d_{t+1} = d_t = 0$ ; clearly no level of aggregation improves the accuracy of forecasting. For a high positive correlated series no level of aggregation may improve the accuracy of forecasts.



**Figure 3-12:** Impact of  $m, \phi, \alpha$  and  $\beta$  on the MSE ratio for  $\alpha = 0.1$  (top) and  $\alpha = 0.5$  (bottom)



**Figure 3-13:** Impact of control parameters for ARIMA(1,0,0) process on the MSE ratio

Hence, when the non-aggregate demand follows an ARIMA(1,0,0) process, the aggregation approach may lead to an improvement in accuracy when the aggregation level,  $m$ , is high and the smoothing constant after aggregation  $\beta$  is small. However, for highly positive values of the autoregressive parameter  $\phi$ , the aggregation approach is not recommended (especially when  $\beta$  is bigger than  $\alpha$ ).

What may be concluded at the end of this sub-section is that if the demand data is highly positive autocorrelated then the non-aggregation approach works better than the aggregation one. In those cases the non-aggregation approach better exploits the very important recent information (i.e.  $d_t$ ) (though it is more prone to noise). On the contrary, when the autocorrelation is less positive or negative, then the recent demand information is not that crucial. Thus, a longer term view of the demand is preferable (if one properly selects how to use long term demand information through  $m$  and  $\beta$ ). Moreover, the aggregation performance under the ARIMA(0,0,1) is slightly different than the ARIMA(1,0,1) and the ARIMA(1,0,0) due to the nature of these processes. Positive autocorrelation under an ARIMA(1,0,1) or ARIMA(1,0,0) process, with a maximum value equal to +1, is potentially higher than that

associated with an ARIMA(0,0,1) process (with a maximum value equal to 0.5). It should be reiterated that for the ARIMA(0,0,1) process, the autocorrelation is limited only for to lag1, whereas for the ARIMA(1,0,1) and the ARIMA(1,0,0) processes, the autocorrelation spans over more lags and is not limited to lag1. This renders the range of outperformance of the non-aggregation approach larger under the ARIMA(1,0,0) and ARIMA(1,0,1) processes. In sub-section 3.2.2 the conditions under which each approach outperforms the other one are theoretically determined when comparison is undertaken at the disaggregate level.

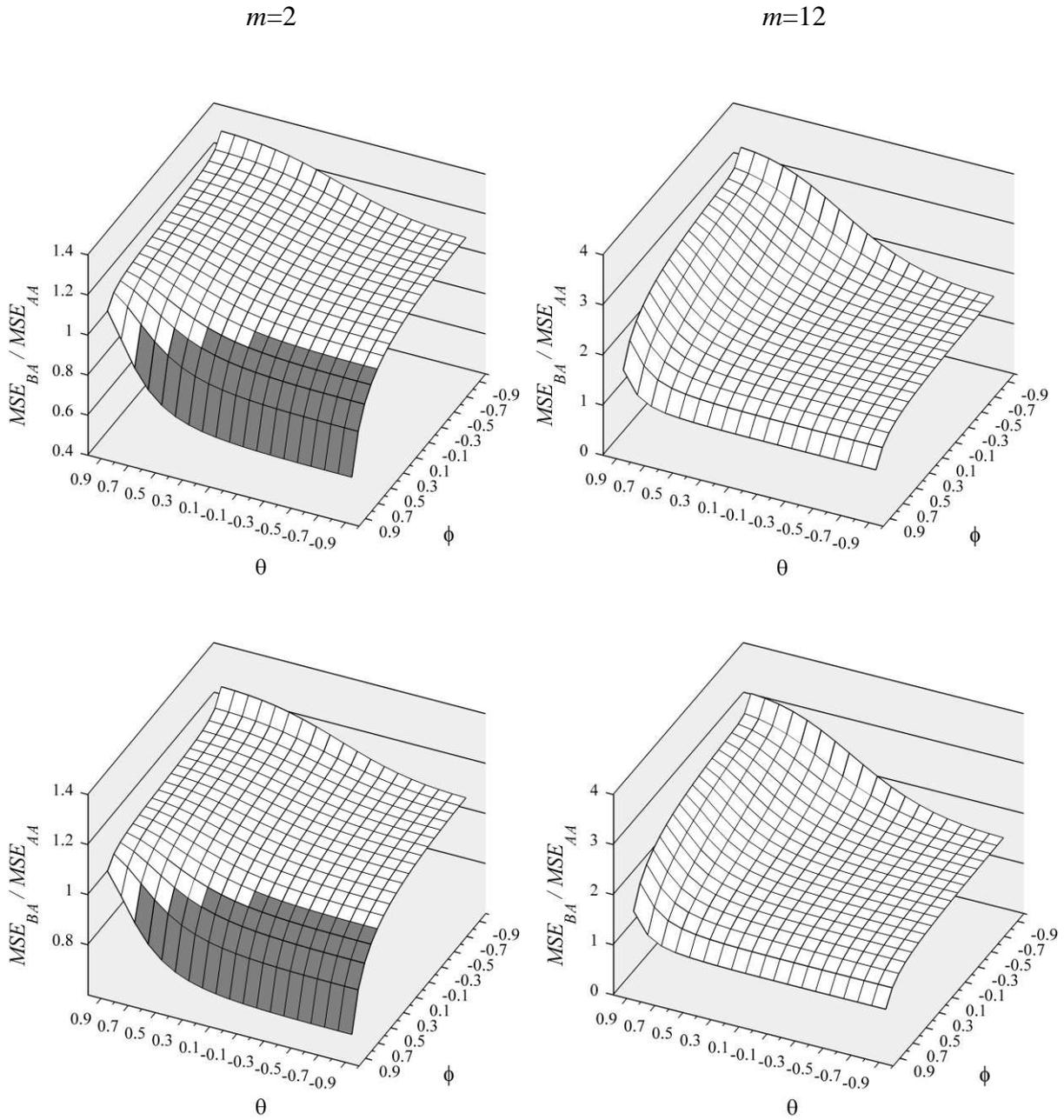
### 3.2.1.2 Comparison at aggregate level

In this sub-section, the effect of control and process parameters on the ratio of  $MSE_{BA}/MSE_{AA}$  is evaluated when the comparison is undertaken at the aggregate level.

#### 3.2.1.2.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

**Figure 3-14** and **Figure 3-15** present the impact of the control and the process parameters on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$ ,  $\alpha = 0.1, \beta = 0.01$  and  $\alpha = 0.1, \beta = 0.05$  when the non-aggregate demand series follows an ARIMA(1,0,1) process.

Shaded areas represent a behavior in favor of the non-aggregation approach. These figures show that the aggregation approach provides more accurate results when the forecast horizon is long. Moreover, for short horizons, the aggregation approach performs extremely well when  $\theta$  is positive and  $\phi$  takes negative values. Alternatively, the aggregation approach does not perform better than the non-aggregation one where  $\theta$  takes negative values and  $\phi$  takes highly positive values. The outperformance of the non-aggregation approach can be attributed to the high positive autocorrelation value as explained above. The results show that the effect of the smoothing constant values before and after aggregation on the superiority of each approach is similar to the case of comparing at the disaggregate level. When considering a fixed value of the aggregation level, increasing  $\beta$  results in a deterioration of the aggregation approach. For the aggregate data, the responsiveness of the stable forecasting method deteriorates the performance because the differences between the observations are small and low  $\beta$  leads to better forecasts.

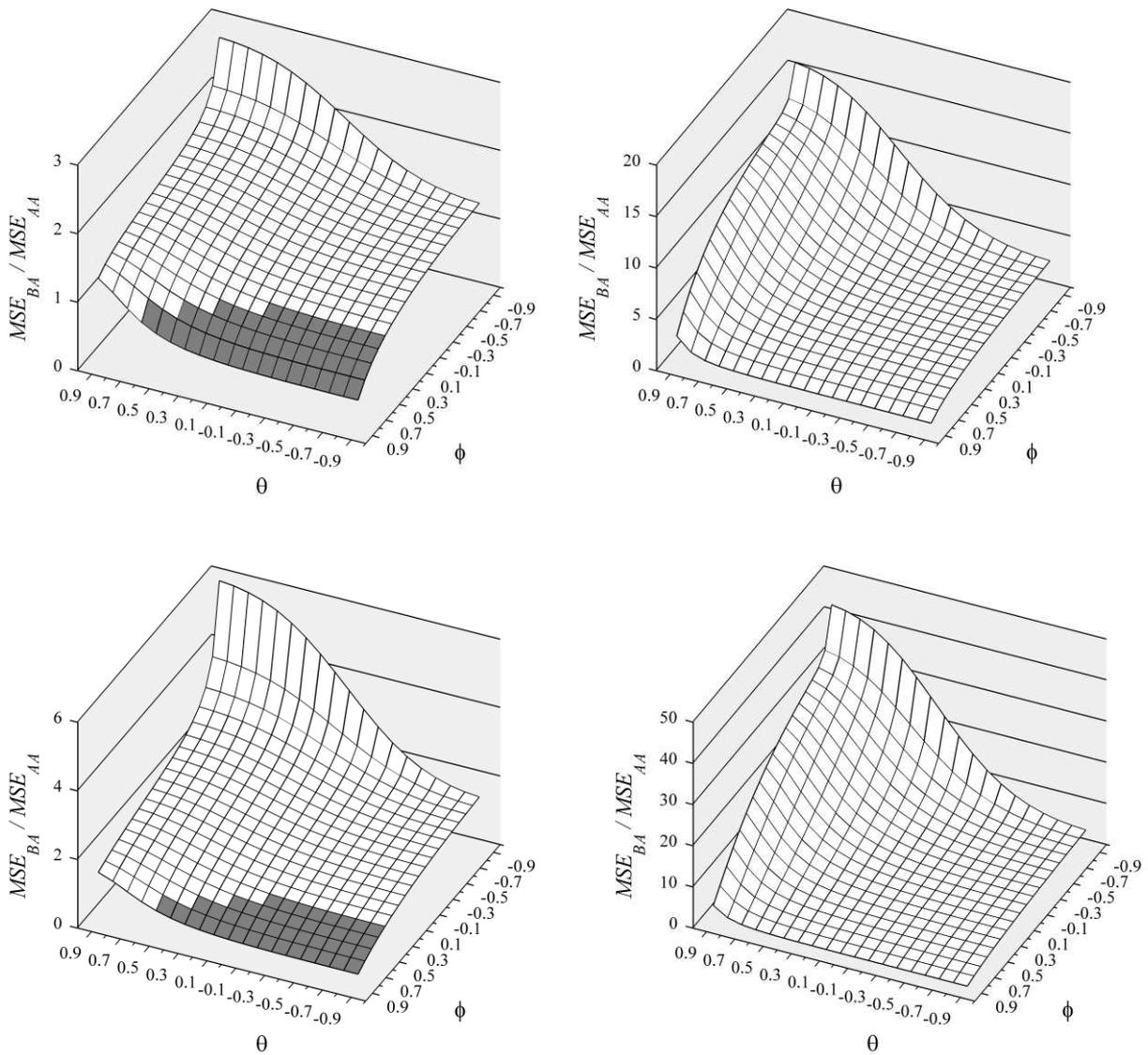


**Figure 3-14:** Impact of  $m$ ,  $\theta$ ,  $\phi$ ,  $\alpha$  and  $\beta$  on the ratio of MSE:

$\alpha = 0.1, \beta = 0.01$  (top)  $\alpha = 0.1, \beta = 0.05$  (bottom)

$m=2$

$m=12$



**Figure 3-15:** Impact of  $m$ ,  $\theta$ ,  $\phi$ ,  $\alpha$  and  $\beta$  on the ratio of MSE:  $\alpha = 0.3, \beta = 0.1$ (top)  
 $\alpha = 0.5, \beta = 0.2$ (bottom)

The results show that by increasing the aggregation level,  $m$  the performance of the aggregation approach is improved. for higher values of the aggregation level, the aggregation approach always outperforms the non-aggregation one regardless of the values of the moving average and the autoregressive parameters. Whereas, when the comparison is considered at the

disaggregate level, for highly positive autocorrelation, no level of aggregation improves the forecast accuracy.

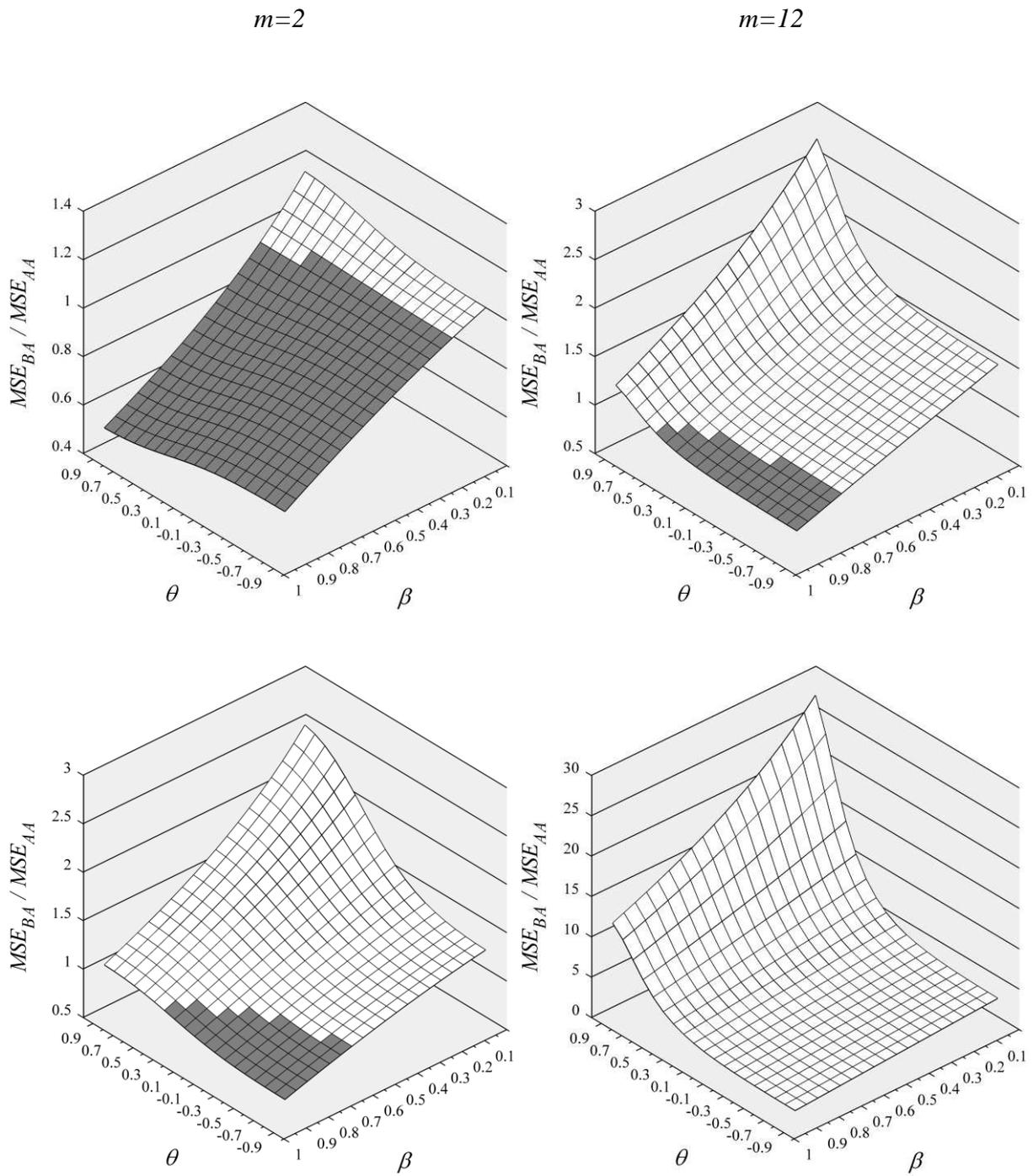
the farther into the future the estimation is calculated, the forecast errors associated with the original data become larger compared to the temporally aggregate one. The approaches based on the temporally aggregate data benefit more by increasing the forecast horizon. In these cases a longer term view on demand becomes vital and the aggregation approach utilizes this information much better than the non-aggregation one.

In the particular case where the smoothing constant parameters before and after aggregation are identical ( $\alpha = \beta$ ), the results are similar to **Figure 3-14** and **Figure 3-15**.

These examples show that the performance superiority of each approach is a function of all the control and the process parameters. The selection of the control parameters  $\alpha$ ,  $\beta$  and  $m$ , influence the effectiveness of the aggregation approach in conjunction with the the process parameters.

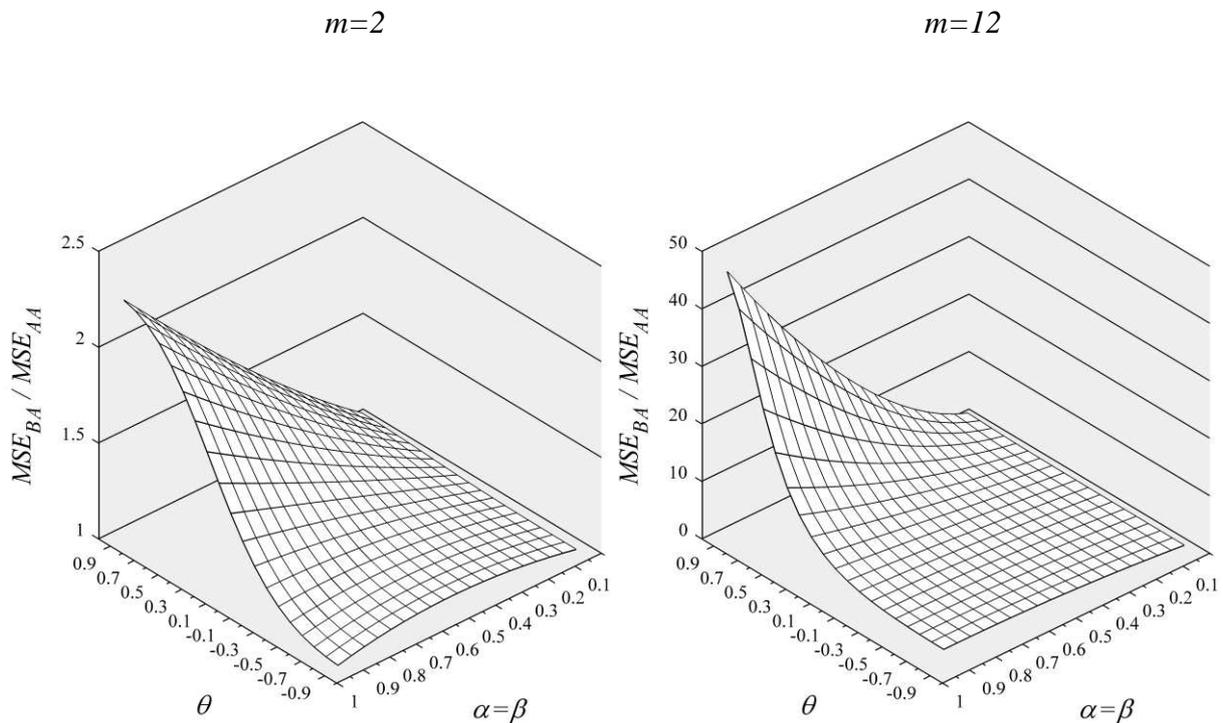
#### 3.2.1.2.2 Moving average process order one, ARIMA(0,0,1)

**Figure 3-16** presents the impact of the control parameters on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$  and  $\alpha = 0.1, 0.5$  when the non-aggregate demand series follows an ARIMA(0,0,1) process. Shaded areas represent a behavior in favor of the non-aggregation approach. The results show that for a fixed value of  $\alpha$ , by increasing the aggregation level  $m$ , the aggregation approach provides more accurate forecasts than the non-aggregation one. On the other hand, when considering a fixed value of the aggregation level, increasing  $\beta$  results in a deterioration of the aggregation approach. If the selected smoothing constant value after aggregation,  $\beta$ , is considerably higher than the smoothing constant used with the original data,  $\alpha$ , then the aggregation approach is not preferred. Alternatively, the aggregation approach yields a more accurate forecast. However, when  $\theta$  takes highly negative values the benefits of the aggregation approach is not as much as positive values. It is obvious from **Figure 3-16** and **Figure 3-17** that there is always a value of  $\beta$  for which the aggregation approach outperforms the non-aggregation one.



**Figure 3-16:** Impact of  $m$ ,  $\theta$ ,  $\alpha$  and  $\beta$  on the MSE ratio for  $\alpha = 0.1$  (top) and  $\alpha = 0.5$  (bottom)

In the particular case where the smoothing constant parameters before and after aggregation are identical (i.e.  $\alpha = \beta$ ), the aggregation approach outperforms the non-aggregation one in all cases.



**Figure 3-17:** Impact of control parameters for ARIMA(0,0,1) process on the MSE ratio

The impact of the smoothing parameter  $\beta$ , and the aggregation level  $m$ , is quite intuitive and similar to the case of the ARIMA(1,0,1) process. Therefore, setting  $\beta$  to be small ( $\beta$  should be smaller than  $\alpha$ ) in conjunction with high aggregation levels provides an advantage to the aggregation approach. This is confirmed by the results presented in **Figure 3-16** and **Figure 3-17**.

The weakness of the aggregation approach for negative values of  $\theta$  can be attributed to the potentially high positive autocorrelation between demand periods. For negative values of  $\theta$ , the autocorrelation is positive; for positive values of  $\theta$  the autocorrelation is negative and for the white noise process, the autocorrelation is zero. Aggregation of a highly positively correlated series doesn't add as much value as the aggregate series with less positive

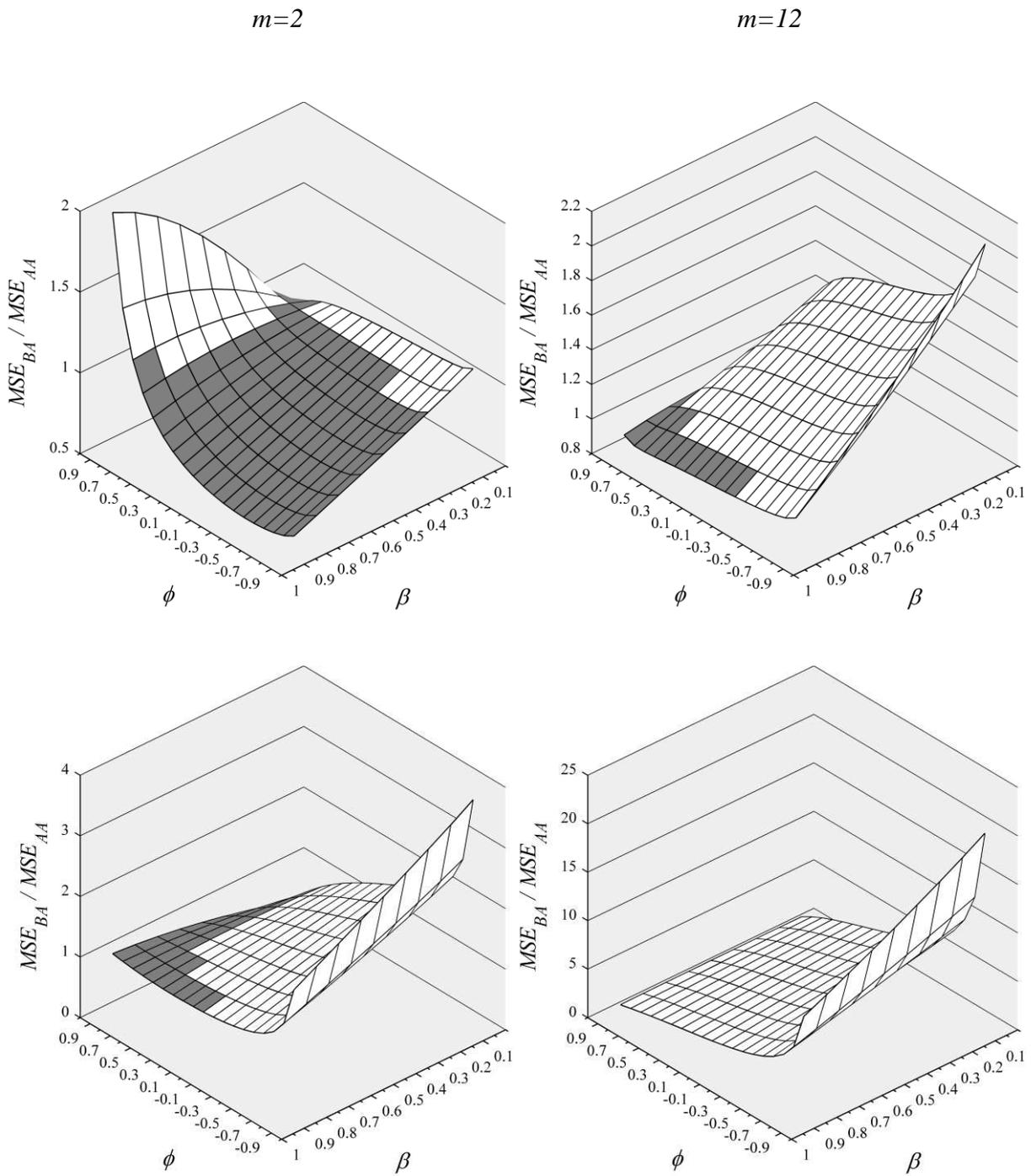
autocorrelation. Hence, when the non-aggregate demand follows an ARIMA(1,0,0) process, the aggregation approach leads to an improvement in accuracy when the aggregation level,  $m$ , is high and the smoothing constant after aggregation  $\beta$  is small.

### 3.2.1.2.3 Autoregressive process order one, ARIMA(1,0,0)

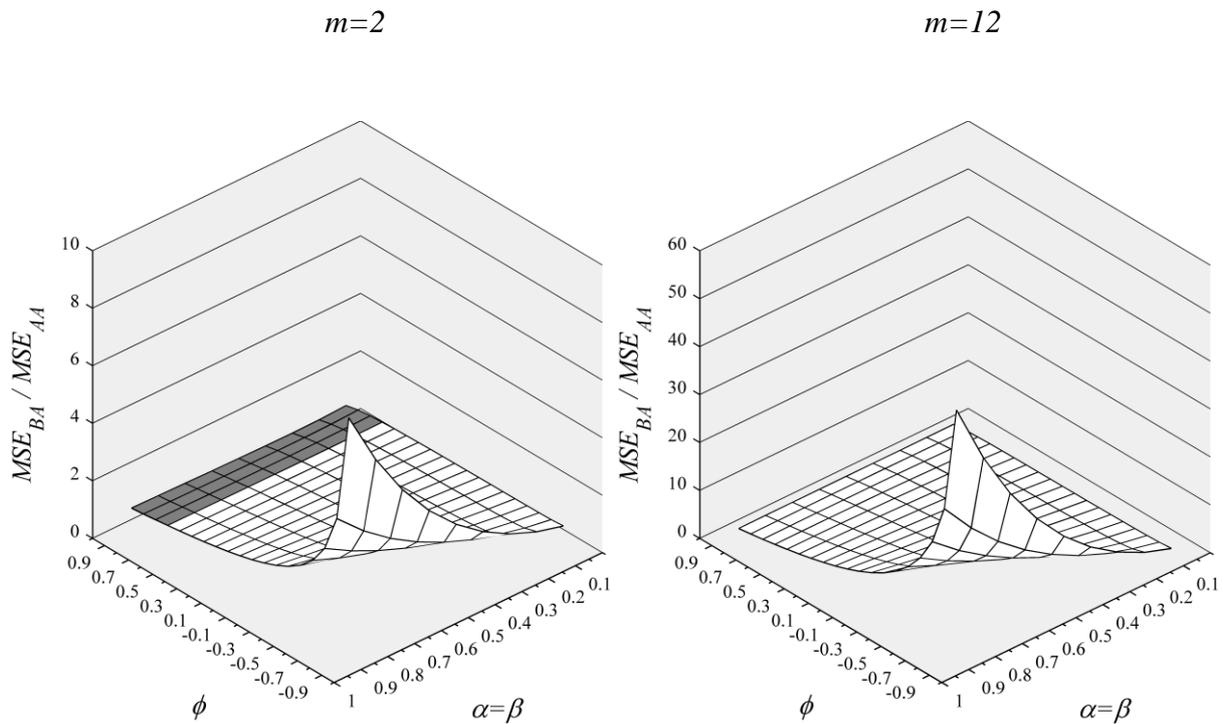
**Figure 3-18** presents the impact of the control parameters  $m$ ,  $\alpha$ ,  $\beta$  on the ratio of  $MSE_{BA}/MSE_{AA}$  for  $m = 2, 12$  and  $\alpha = 0.1, 0.5$ , when the non-aggregate demand series follows an ARIMA(1,0,0) process. It is easy to see that the superiority of each approach is a function of all control and process parameters.

The results show that for a fixed value of  $\alpha$ , increasing the aggregation level results in accuracy improvements of the aggregation approach. Conversely, for a fixed aggregation level, increasing  $\beta$  results in a deterioration of the performance. In addition,  $\beta$  should be generally smaller than  $\alpha$  in order for the aggregation approach to produce more accurate forecasts.

When the smoothing constant parameters before and after aggregation are identical (i.e.  $\alpha = \beta$ ), the aggregation approach outperforms the non-aggregation one in all cases except when the aggregation level is low and associated with highly positive values of  $\phi$ . Moreover, by increasing the aggregation level the performance of the aggregation approach is improved and for the higher aggregation level, the aggregation approach always performs better.



**Figure 3-18:** Impact of  $m$ ,  $\phi$ ,  $\alpha$  and  $\beta$  on the MSE ratio for  $\alpha = 0.1$  (top) and  $\alpha = 0.5$  (bottom)



**Figure 3-19:** Impact of control parameters for ARIMA(1,0,0) process on the MSE ratio

The impact of the smoothing parameter  $\beta$  and the aggregation level  $m$  on the ratio is similar to those reported for the ARIMA(0,0,1) process. Figure 3-18 shows that even if the selected  $\beta$  is smaller than  $\alpha$ , there are some cases in which the aggregation approach is not preferred. This is when the autoregressive parameter  $\phi$  takes high positive values. In general, the benefits achieved by the aggregation approach are fewer for highly positive values of  $\phi$  than negative values of  $\phi$ . Hence, when the non-aggregate demand follows an ARIMA(1,0,0) process, the aggregation approach leads to an improvement in accuracy when the aggregation level,  $m$ , is high and the smoothing constant after aggregation  $\beta$  is small.

What can be concluded at the end of this sub-section in forecasting the aggregate level is that if the forecast horizon is long then the aggregation approach is always preferred. Because in these cases a longer term view on demand is very important and the aggregation approach utilizes this information better than the non-aggregation one. By increasing the

forecast horizon, the forecast error associated with the classical approach increases as well. However, when the forecast horizon is short, the superiority of each approach depends on the aggregation level and the autocorrelation values. If the demand data is highly positive autocorrelated then the non-aggregation approach works better than the aggregation one. In those cases the non-aggregation approach better exploits the very important recent information (i.e.  $d_t$ ). On the contrary, when the autocorrelation is less positive or negative, the recent demand information is not as crucial. Thus, a longer term view of the demand is becomes important. Therefore, the aggregation approach is preferred.

Moreover, the aggregation performance under the ARIMA(1,0,1), the ARIMA(0,0,1) and the ARIMA(1,0,0) processes is slightly different due to the nature of these processes. In fact, the positive autocorrelation in the the ARIMA(1,0,1) and the ARIMA(1,0,0) is higher than that in the ARIMA(0,0,1) which makes larger the range of the outperformance of the non-aggregation approach in the the ARIMA(1,0,1)and the ARIMA(1,0,0) processes.

In sub-section 3.2.2 the conditions under which each approach outperforms the other one are theoretically determined when comparison is undertaken at the aggregate level.

### **3.2.2 Theoretical Comparison**

Having conducted a sensitivity analysis in sub-section 3.2.1, now the conditions under which each approach outperforms the other one are analytically identified.

#### **3.2.2.1 Comparison at disaggregate level**

In this sub-section the conditions under which the aggregation and the non-aggregation approaches perform better are identified when the comparison is undertaken at the disaggregate level.

### 3.2.2.1.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

In this sub-section the conditions under which each approach outperforms the other one are analytically identified when the non-aggregate demand process is an ARIMA(1,0,1). The ratio of  $MSE_{BA}/MSE_{AA}$  is obtained by dividing (3-17) into (3-28) :

$$MSE_{BA}/MSE_{AA} = \frac{1}{1-0.5\alpha} \left( \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} - \frac{\alpha \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right)}{1-\phi + \alpha\phi} \right) \left( \frac{\left( \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} \right) + \frac{1}{m^2} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right)}{2-\beta} + \frac{2\beta(1-\beta) \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \alpha\phi^m)} \right) - \frac{2}{m} \left( \frac{\beta \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right)}{1-\phi^m + \alpha\phi^m} \right) \times \frac{1-\phi^m}{1-\phi} \quad (3-49)$$

This ratio is a function of the aggregation level  $m$ , the autoregressive parameter  $\phi$ , moving average parameter  $\theta$ , and the smoothing constant parameters before and after aggregation,  $\alpha$  and  $\beta$ . Considering that the aggregation level may only get integer values greater than or equal to two, the goal is to determine the value  $\beta$  that enables the aggregation approach to perform better. The entire range of possible values for  $\alpha$  is considered.

To show the conditions under which the aggregation approach outperforms the non-aggregation one, the equation (3-49) is set greater than 1, i.e.  $MSE_{BA}/MSE_{AA} > 1$ , From this statement the following result can be obtained:

**THEOREM 1-3:** If the time series of the non-aggregate demand follows an ARIMA(1,0,1) process and  $-1 < \theta < 1$  and  $-1 < \phi \leq 0$ , then:

- If  $\beta < \beta_1$ , the aggregation approach provides more accurate forecast.
- If  $\beta = \beta_1$ , both approaches perform equally.
- Otherwise, the non-aggregation approach works better.

Where  $\beta_1$  defined in (F-4).

**PROOF:** the proof of Theorem 1-3 is given in Appendix F.

Note that for the presented range of  $\theta$  and  $\phi$  ( $-1 < \theta < 1$  and  $-1 < \phi \leq 0$ )  $\beta_1$  is always positive, consequently choosing  $\beta < \beta_1$  guarantees that the aggregation approach always outperforms the non-aggregation one in this region. Hence, the value of  $\beta_1$  reflects a cut-off point that may be used in practice for the selection of the smoothing constant value to be used for the aggregate series. The cut-off point reflects all the qualitative discussion provided in the previous sub-section as to when aggregation outperforms the non-aggregation approach.

If the time series of the original demand follows an ARIMA(1,0,1) process and the moving average and the autoregressive parameters satisfy  $-1 < \theta < 1$  and  $0 < \phi < 1$ , then the conditions under which each approach works better can be obtained. These conditions are summarized in the following selection procedure (discussed in Table 1):

**Table 3-2:** Selection procedure for the ARIMA(1,0,1) process, Comparison at disaggregate level

- 
1. The procedure is begun by calculating  $\Delta$  defined in (F-3), If  $\Delta < 0$  then the non-aggregate approach is always superior, otherwise the values of  $\beta_1$  and  $\beta_2$  defined in (F-4) and (F-5) are calculated.
  2. If  $\beta_2 \in (0,1)$ , the value of  $\beta_1$  and according to the values of  $\beta_1$  and  $\beta_2$  the following rules are obtained:
    - If  $\beta_2 < \beta < \beta_1$ , then the aggregation approach works better.
    - If  $\beta = \beta_1 = \beta_2$  then both approaches are identical.
    - If  $\beta > \beta_1$  or  $\beta < \beta_2$  then non-aggregate strategy works better. Otherwise, go to 3.
-

- 
3. If  $\beta_2 \notin (0,1)$ , the value of  $\beta_1$  is calculated:
- If  $\beta < \beta_1$ , then the aggregation approach works better.
  - If  $\beta = \beta_1$ , then both approaches are identical.
  - If  $\beta > \beta_1$ , then nonaggregation approach works better.
- 

Where  $\beta_2$  defined in (F-5).

**PROOF:** The details of the selection procedure are given in Appendix G.

### 3.2.2.1.2 Moving average process order one, ARIMA(0,0,1)

The ratio of the  $MSE_{BA}$  to  $MSE_{AA}$  when the non-aggregate demand follows an ARIMA(0,0,1) process is a function of the moving average parameter, the smoothing constant before and after aggregation ( $\alpha$  and  $\beta$ ), and the aggregation level. The cut-off points for the value of  $\beta$  should be determined. This enables the aggregation approach to perform better. The entire range of possible values for  $\alpha$  is considered but the smoothing constant is a parameter that is set to its optimal value by practitioners, normally by minimizing the  $MSE$ . From (3-17) it is clear that  $MSE_{BA}$  is monotonically increasing in  $\alpha$  as the derivative of  $MSE_{BA}$  is positive for all values of  $\theta$  in  $(-1, 1)$ . Hence,  $MSE_{BA}$  can be minimized by having the smallest possible value of  $\alpha$ , which makes sense for a stationary process. However, it should be noted that in this theoretical analysis the issue of initialization of the forecasting process is disregarded. This is an important issue to be mentioned (since with very low  $\alpha$  values a bad initialization implies inaccurate estimates of the future demand as the forecast will basically be kept constant) but one that is not considered as part of this research.

To show the conditions under which the aggregation approach outperforms the non-aggregation approach, the ratio is set greater than one,  $MSE_{BA}/MSE_{AA} > 1$ . From this inequality the following result can be obtained:

**THEOREM 2-3:** If the time series of the non-aggregate demand follows an ARIMA(0,0,1) process, then:

- If  $\beta < \beta_1$ , the aggregation approach provides more accurate forecasts.
- If  $\beta = \beta_1$ , both strategies perform equally.
- Otherwise, the non-aggregation approach works better.

where

$$\beta_1 = \frac{\left( -(m^2\eta + m(1 + \theta^2) + 2m\theta) + \sqrt{\left( (m^2\eta + m(1 + \theta^2) + 2m\theta) \right)^2 + 8(2\theta - 2m\theta)m^2\eta} \right)}{2(2\theta - 2m\theta)}, \quad (3-50)$$

and

$$\eta = \frac{\alpha(1 + \theta)^2}{2 - \alpha}. \quad (3-51)$$

**PROOF:** the proof of Theorem 2-3 is given in Appendix H.

The results demonstrate that, for a given values of  $\alpha$  and  $m$ , there always exists a value of  $\beta$  such that the aggregation approach outperforms the non-aggregation one. Hence, the value of  $\beta_1$  reflects a cut-off point that may be used in practice for the selection of the smoothing constant value to be used for the aggregate series.

### 3.2.2.1.3 Autoregressive process order one, ARIMA(1,0,0)

A similar procedure is followed by setting the ratio  $MSE_{BA}$  to  $MSE_{AA}$  greater than 1 for an ARIMA(0,0,1) process. This is conducted to identify the conditions under which the

aggregation approach performs better. These conditions are summarized by the selection procedure presented in Appendix I when autoregressive parameter  $\phi$  satisfies  $-1 < \phi \leq +1$ .

As discussed earlier the smoothing constant is often set by practitioners to its optimal value, so it is more interesting to discuss the cases where such a value is considered. To do so, the value that minimizes the  $MSE_{BA}$  is determined. Following that a value of the smoothing constant after aggregation that leads to more accurate forecasts is calculated. The optimal value of  $\alpha$  is given in (3-52) that can be obtained by solving the first derivative of (3-19):

$$\alpha^* = \begin{cases} (3\phi - 1)/2\phi & 1/3 < \phi < 1 \\ \omega & -1 < \phi \leq 1/3 \end{cases} \quad (3-52)$$

where  $\omega > 0$  is a very small positive value.

By considering the optimal value of the smoothing constant before aggregation, two different cases should be considered. From  $MSE_{BA}/MSE_{AA} > 1$  and (3-52) the following results can be obtained.

**Case 1.**  $1/3 < \phi < 1$ . In this case,  $\alpha^* = (3\phi - 1)/2\phi$

**THEOREM 3-3:** If the time series of the non-aggregate demand follows an ARIMA(1,0,0) process, where  $1/3 < \phi < 1$  and the optimal smoothing constant,  $\alpha^* = (3\phi - 1)/2\phi$ , is used to determine the non-aggregate demand forecast, then the non-aggregation approach always provides more accurate forecast than the aggregation one, regardless of the smoothing constant parameter after aggregation,  $\beta$ , and the aggregation level,  $m$ .

**PROOF:** the proof of Theorem 3-3 is given in Appendix J.

**Case 2.**  $-1 < \phi \leq 1/3$ . In this case  $\alpha^*$  is a very small positive number.

**THEOREM 4-3.** If the time series of the non-aggregate demand follows an ARIMA(1,0,0) process, where  $-1 < \phi \leq 1/3$  and the optimal smoothing constant used to determine the non-aggregate demand forecast,  $\alpha^* < 0.05$ , then:

- If  $\beta < \beta_1$  the aggregation approach provides more accurate forecast.
- If  $\beta = \beta_1$  both strategies perform equally.
- Otherwise, the non-aggregation approach works better.

Where

$$\beta_1 = \frac{-((1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)) + \sqrt{\Delta}}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}. \quad (3-53)$$

( $\xi_1, \xi_2, \eta$  and  $\Delta$  are given in Appendix I)

**PROOF:** The proof of Theorem 4-3 is given in Appendix J.

Similar to the case of the ARIMA(0,0,1) process, the above results provide a cut-off point that may be used in practice for the selection of the smoothing constant in order to obtain an outperformance of the aggregation approach when ARIMA(1,0,0) processes are considered. Obviously, as the cut-off point increases for high aggregation levels, it is clear that this implies a considerable range of the smoothing constant of the aggregate series where there is a benefit of using the aggregation approach. Hence, these results provide a comprehensive way of managing the process of forecasting of ARIMA(1,0,0) processes when the autoregressive parameter is known and when the intention is to optimize the smoothing constant for the non-aggregate series.

### 3.2.2.2 Comparison at aggregate level

In this sub-section the superiority conditions of each approach are identified when the comparison is undertaken at the aggregate level.

#### 3.2.2.2.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

The ratio of the  $MSE_{BA}$  to  $MSE_{AA}$  when the non-aggregate demand follows an ARIMA(1,0,1) process is a function of the moving average parameter,  $\theta$ , the autoregressive

parameter,  $\phi$ , the smoothing constant before and after aggregation ( $\alpha$  and  $\beta$ ), and the aggregation level,  $m$ . The objective is to determine the value  $\beta$  that enables the aggregation approach to perform better.

$$\frac{MSE_{BA}}{MSE_{AA}} = \left( \frac{\left( \frac{m(1-2\phi\theta + \theta^2)}{1-\phi^2} \right) + \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right)}{2-\alpha \left( \frac{m^2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) - \frac{2m\alpha[(1-\phi)+\phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \right)} \right) \Bigg/ \left( \frac{2 \left( \frac{m(1-2\phi\theta + \theta^2)}{1-\phi^2} \right) + \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right)}{2-\beta} - \frac{2\beta \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} \left( \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \right) \tag{3-54}$$

To show the conditions under which the aggregation approach outperforms the non-aggregation approach, the ratio is set to greater than one,  $MSE_{BA}/MSE_{AA} > 1$ . From this statement the following results can be obtained:

If the time series of the basic demand follows an ARIMA(1,0,1) process and the moving average and the autoregressive parameters satisfy  $-1 < \theta < 1$  and  $0 < \phi < 1$ , the conditions under which each approach works better are obtained. These conditions are summarized as follows:

**Table 3-3:** Selection procedure for the ARIMA(1,0,1) process, Comparison at aggregate level

- 
1. The procedure is begun by calculating  $\Delta$  defined in (K-2), If  $\Delta < 0$  then the non-aggregation approach is always superior, otherwise the values of  $\beta_1$  and  $\beta_2$  defined in (K-3) and (K-4) are calculated.
  2. If  $\beta_2 \in (0,1)$ , the value of  $\beta_1$  and according to the values of  $\beta_1$  and  $\beta_2$  the following rules
-

---

are obtained:

- If  $\beta_2 < \beta < \beta_1$ , then the aggregation approach works better.
  - If  $\beta = \beta_1 = \beta_2$  then both approaches are identical.
  - If  $\beta > \beta_1$  or  $\beta < \beta_2$  then non-aggregation approach works better. Otherwise, go to 3.
3. If  $\beta_2 \notin (0,1)$ , the value of  $\beta_1$  is calculated:
- If  $\beta < \beta_1$ , then the aggregation approach works better.
  - If  $\beta = \beta_1$ , then both approaches are identical.
  - If  $\beta > \beta_1$ , then non-aggregation approach works better.
- 

**PROOF:** The details of the selection procedure are given in Appendix K.

**THEOREM 5-3:** If the time series of the non-aggregate demand follows an ARIMA(1,0,1) process and  $-1 < \theta < 1$  and  $-1 < \phi \leq 0$ , then:

- If  $\beta < \beta_1$ , the aggregation strategy provide more accurate forecast.
- If  $\beta = \beta_1$ , both strategies perform equally.
- Otherwise, the non-aggregation strategy works better.

where  $\beta_1$  is defined as:

$$\beta_1 = \frac{(2\phi^m\Psi - \Psi(1 - \phi^m) - 2\eta\phi^m + 2\xi) + \sqrt{\Delta}}{2\phi^m\Psi}$$

$\Psi, \eta, \xi$ , and  $\Delta$  are defined in Appendix K.

**PROOF:** the proof of Theorem 5-3 is given in Appendix L.

Theorem 5-3 show that when the autoregressive and the moving average parameters satisfies  $-1 < \theta < 1$  and  $-1 < \phi \leq 0$ , then for a given value of the smoothing constant,  $\alpha$ , and the aggregation level,  $m$ , there is always a value of  $\beta$  for which the aggregation approach provides more accurate forecasts.

### 3.2.2.2.2 Moving average process order one, ARIMA(0,0,1)

The ratio of the  $MSE_{BA}$  to  $MSE_{AA}$  when the non-aggregate demand follows an ARIMA(0,0,1) process is a function of the moving average parameter,  $\theta$ , the smoothing constant before and after aggregation ( $\alpha$  and  $\beta$ ), and the aggregation level,  $m$ . The superiority conditions can be obtained by following the same procedure as Appendix K where the autoregressive parameter is equal to zero.

$$MSE_{BA}/MSE_{AA} = \frac{\left( (1+\theta^2) + 2\alpha\theta + m \left( \frac{\alpha(1+\theta^2)}{2-\alpha} - \frac{2\alpha(1-\alpha)\theta}{(2-\alpha)} \right) \right)}{\left( \frac{2(1+\theta^2)}{2-\beta} \right)} \quad (3-55)$$

By setting the equation (3-55) to greater than one, the following results can be obtained:

**THEOREM 6-3:** If the time series of the non-aggregate demand follows an ARIMA(0,0,1) process, then for a given values of  $\alpha$  and  $\theta$ :

- If  $\beta < \beta_1$ , the aggregation approach provides more accurate forecasts.
- If  $\beta = \beta_1$ , both strategies perform equally.
- Otherwise, the non-aggregation approach works better.

where

$$\beta_1 = 2 - \frac{2m(1+\theta^2)}{\left( m(1+\theta^2) + 2m\alpha\theta + \frac{\alpha m^2(1+\theta^2)}{2-\alpha} - \frac{2\alpha m^2(1-\alpha)\theta}{(2-\alpha)} \right)}$$

Proof: Theorem 6-3 can be obtained by substituting  $\phi=0$  in Appendix L.

Theorem 6-3 says that there is always a value of  $\beta$  for which the aggregation approach outperforms the non-aggregation one.

**THEOREM 7-3** If the non-aggregate time series follows an ARIMA(0,0,1) process and the smoothing constant under the aggregation approach is smaller or equal to the non-aggregation one ( $\beta \leq \alpha$ ), then aggregation approach always outperforms the non-aggregation one (i.e.  $MSE_{BA} > MSE_{AA}$ ). This is true regardless of the aggregation level,  $m$  and the process parameter  $\theta$ . In addition, when the smoothing constants under the both approaches are set small ( $\alpha, \beta < 0.01$ ), then both aggregation and non-aggregation approaches perform equally.

**PROOF:** the proof of Theorem 7-3 is given in Appendix M.

### 3.2.2.2.3 Autoregressive process order one, ARIMA(1,0,0)

The superiority conditions of each approach when the non-aggregate demand follows an ARIMA(1,0,0) process can be obtained by setting the following equation greater than one.

$$MSE_{BA} / MSE_{AA} = \frac{\left[ m + \phi \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) + \frac{m^2 \alpha}{(2-\alpha)} + \frac{\phi m^2 2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)} - \frac{2m\alpha\phi[(1-\phi)+\phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)} \right]}{\left[ \frac{2 \left( m + \phi \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right)}{2-\beta} - \frac{2\beta\phi \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m + \beta\phi^m)} \right]} \quad (3-56)$$

Similar to the case of comparison at disaggregate level, by considering the optimal value of the smoothing constant before aggregation  $\alpha$ , two different cases are considered.

**THEOREM 8-3** If the time series of the non-aggregate demand follows an ARIMA(1,0,0) process when  $-1 < \phi \leq 1/3$ , then there is always a value of  $\beta$  in order to the aggregation approach outperforms the non-aggregation one:

- If  $\beta < \beta_1$  the aggregation approach provides more accurate forecast.
- If  $\beta = \beta_1$  both strategies perform equally.
- Otherwise, the non-aggregate approach works better.

Where

$$\beta_1 = \frac{(2\phi^m\Psi - \Psi(1 - \phi^m) - 2\eta\phi^m + 2\xi) + \sqrt{\Delta}}{2\phi^m\Psi}$$

Proof: These conditions can be achieved by substituting  $\theta=0$  and  $-1 < \phi \leq 1/3$  in the Appendix K.

If the autoregressive parameter satisfies  $-1 < \phi \leq 1/3$ , then there is always a value of  $\beta$  for which the aggregation approach works better than the non-aggregation one.

**THEOREM 9-3** If the time series of the non-aggregate demand follows an ARIMA(1,0,0) process when  $-1 < \phi \leq 1/3$  and the smoothing constant under the aggregation approach is smaller than non-aggregation one ( $\beta < \alpha$ ), then aggregation approach always outperforms the non-aggregation one. This is always true regardless of the aggregation level,  $m$ . In addition, when the smoothing constants under the both approaches are set to small values ( $\alpha, \beta < 0.01$ ), then the difference in the performance of the aggregation and non-aggregation approaches is insignificant.

**PROOF:** the proof of Theorem 9-3 is given in Appendix N.

If the time series of the non-aggregate demand follows an ARIMA(1,0,0) process when  $1/3 < \phi \leq 1$ , then the ratio of  $MSE_{BA}/MSE_{AA}$  may be smaller, greater than or equal to one

depending on the values of the smoothing constants ( $\alpha$  and  $\beta$ ), aggregation level,  $m$  and the autoregressive parameters. The conditions under which each approach works better can be obtained by substituting  $\theta=0$  and  $1/3 < \phi \leq 1$  in the procedure discussed in sub-section 3.2.2.2.1 where the case of the ARIMA(1,0,1) is considered.

### 3.2.3 Optimal aggregation level

The objective of this section is to identify the optimal aggregation levels that maximize the ratio or equivalently minimize the  $MSE_{AA}$  for each demand process under consideration. To do so, the ratio of  $MSE_{BA}$  to  $MSE_{AA}$  for the whole range of the control parameters is evaluated.

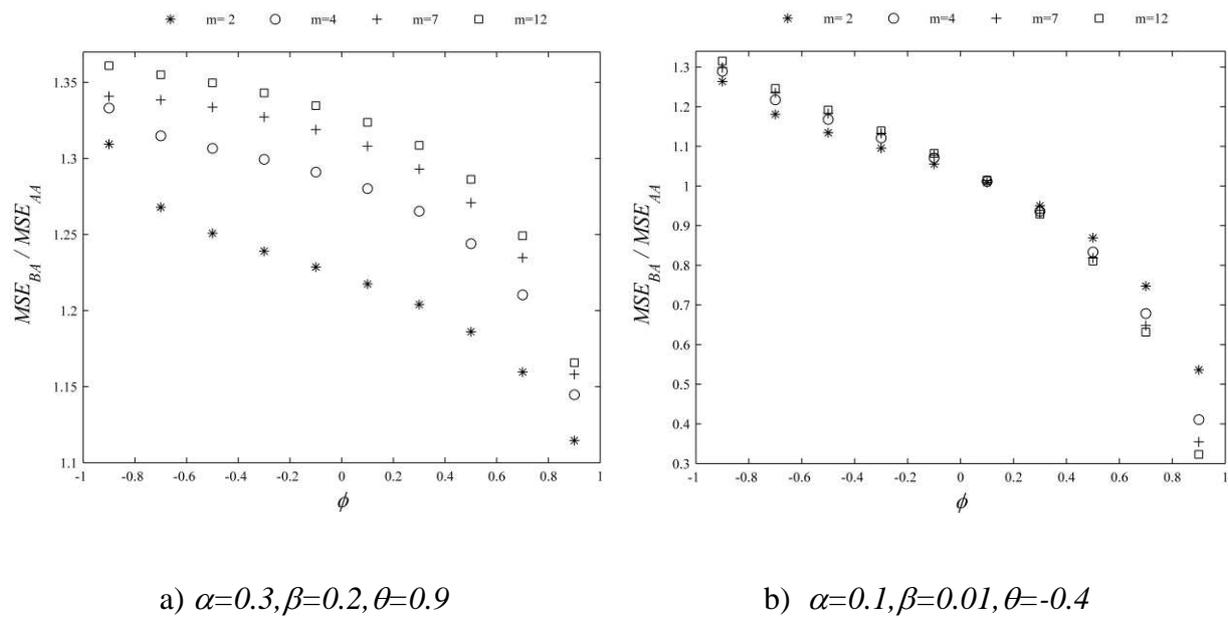
#### 3.2.3.1 Comparison at disaggregate level

In this part the aggregation level that leads to more error reduction is determined when the comparison is undertaken at disaggregate level.

##### 3.2.3.1.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

A numerical investigation to determine the optimal aggregation level is conducted since from (3-49) it is clear that the calculation of the first derivative is infeasible. Two examples are presented: i) the whole range of  $\phi$  where  $\theta=0.9$ ,  $\alpha = 0.3$ , and  $\beta = 0.2$ ; ii) the whole range of  $\phi$  where  $\theta=-0.5$ ,  $\alpha = 0.3$ , and  $\beta = 0.2$ . In the latter case for some values of  $\phi$  and  $\theta=-0.5$  ( **Figure 3-20b**) the ratio is smaller than one and consequently aggregation does not work. Thus, in these cases it is not necessary to discuss the optimal aggregation level.

The results show that by increasing the aggregation level, the performance of the aggregation approach improves. Additionally, a higher aggregation level results in higher values of the ratio and consequently more benefits for the aggregation approach.



**Figure 3-20:** MSE ratio for different values of m for an ARIMA(1,0,1) process

**3.2.3.1.2 Moving average process order one, ARIMA(0,0,1)**

In order to obtain the optimal aggregation level when the non-aggregate demand series follows an ARIMA(0,0,1), the following theorem is considered.

**THEOREM 10-3:** If the non-aggregate demand series follows an ARIMA(0,0,1) process, then the optimal aggregation level is the highest level in any considered range.

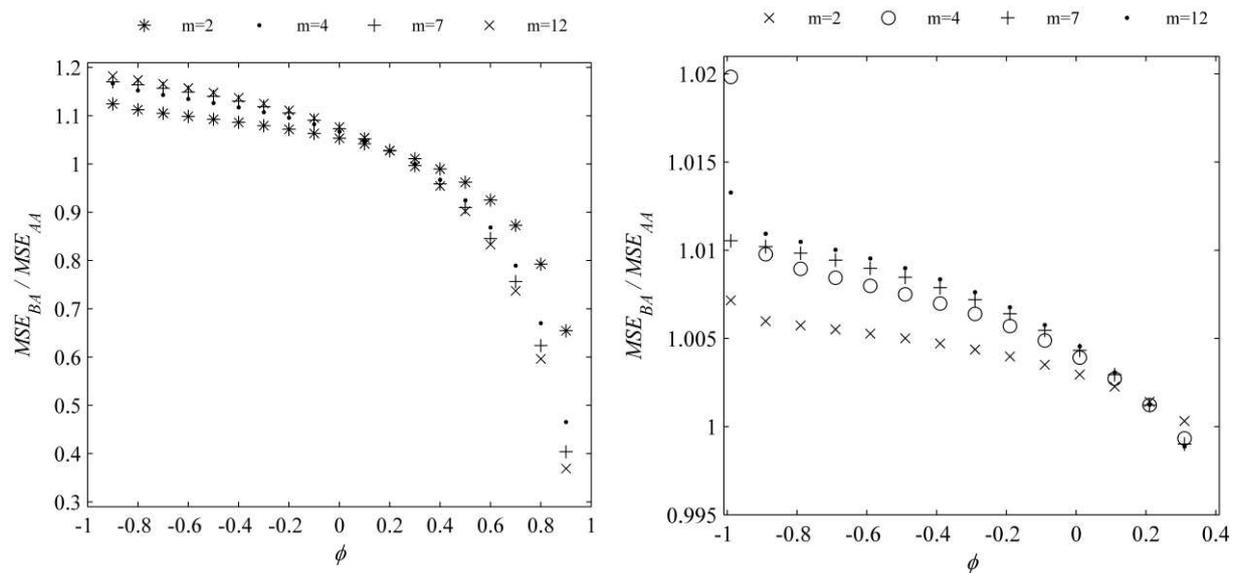
Supposing that aggregation is to be tested in a range  $[u_1, u_2]$ , where  $u_1$  and  $u_2$  are the lower and upper bound, respectively. In addition, they are positive integer numbers. The optimal aggregation level is always  $u_2$ .

**PROOF:** A calculation of the first derivative of  $MSE_{AA}$  with respect to  $m$  shows that  $MSE_{AA}$  is a decreasing function of  $m$ . This can be shown by a numerical analysis for  $m \geq 2$  as well. This means that the ratio  $MSE_{BA}/MSE_{AA}$  is an increasing function of  $m$ . Therefore, a higher value of the aggregation level results in a higher value of the ratio  $MSE_{BA} / MSE_{AA}$ .

**3.2.3.1.3 Autoregressive process order one, ARIMA(1,0,0)**

A numerical investigation is conducted to obtain the optimal aggregation level where the subaggregate process follows an ARIMA(1,0,0) as the calculation of the first derivative is infeasible. Two examples are considered: i) the whole range of  $\phi$  where  $\alpha = 0.15$  and  $\beta = 0.1$ ; ii) the case 2 discussed in 5.2.2 with an optimal value of  $\alpha$ .

Figure 3a shows that the value of the aggregation level that maximizes the MSE ratio changes when varying the control parameter values. For negative and lower positive values of  $\phi$ , i.e.  $-1 < \phi \leq 1/3$ , the forecast accuracy of the aggregation approach increases with the aggregation level while for higher positive values of  $\phi$ , i.e.  $1/3 < \phi < 1$ , this is not true. Let us analyse the two different cases in which the optimal smoothing constant values are considered for  $MSE_{BA}$ .



a) ARIMA(1,0,0) process where  $-1 < \phi \leq 1$

b) Case 2, ARIMA(1,0,0) process where  $-1 < \phi \leq 0.33$

**Figure 3-21:** MSE ratio for different values of m ARIMA(1,0,0) process at disaggregate level

**Case 1.**  $1/3 < \phi < 1$ . In this case the optimal smoothing constant parameter  $\alpha^* = (3\phi - 1)/2\phi$  is used and it is seen in sub-section 5.2.2 that the *MSE* ratio is always lower than 1.

**Case 2.**  $-1 < \phi \leq 1/3$ . In this case a very small smoothing constant value,  $\alpha^* \leq 0.05$ , is used. The *MSE* ratio for different aggregation levels is shown in **Figure 3-21b** for  $-1 < \phi \leq 1/3$  and a numerical example of  $\alpha$  and  $\beta$  values where  $\beta < \alpha$ . This figure shows that the aggregation approach is associated with more accurate results for higher aggregation levels.

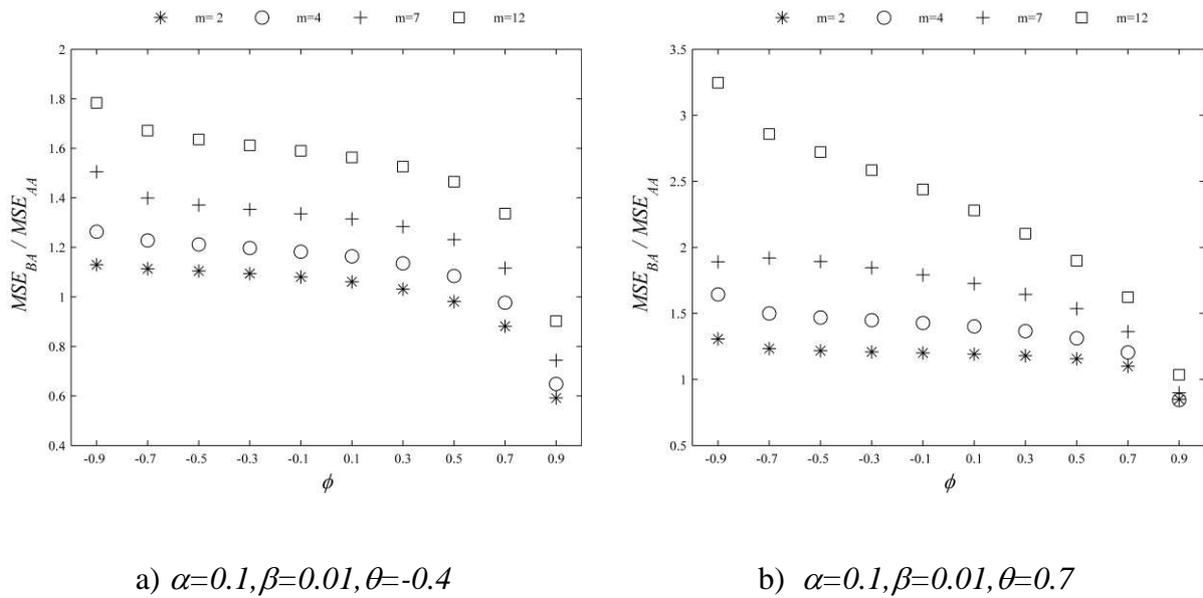
### 3.2.3.2 Comparison at aggregate level

In this part the optimal aggregation level that maximizes the ratio of  $MSE_{BA} / MSE_{AA}$  is identified when the comparison is undertaken at the aggregate level.

#### 3.2.3.2.1 Autoregressive Moving Average Process Order One, ARIMA(1,0,1)

Similar to the case of comparison at disaggregate level, here two examples are presented to evaluate the impact of the aggregation level on the ratio as the derivation of (3-54) to determine the optimal  $m$  is not feasible: i) the whole range of  $\phi$  where  $\theta=0.7$ ,  $\alpha = 0.1$ , and  $\beta = 0.01$ ; ii) the whole range of  $\phi$  where  $\theta=-0.4$ ,  $\alpha = 0.1$ , and  $\beta = 0.01$ .

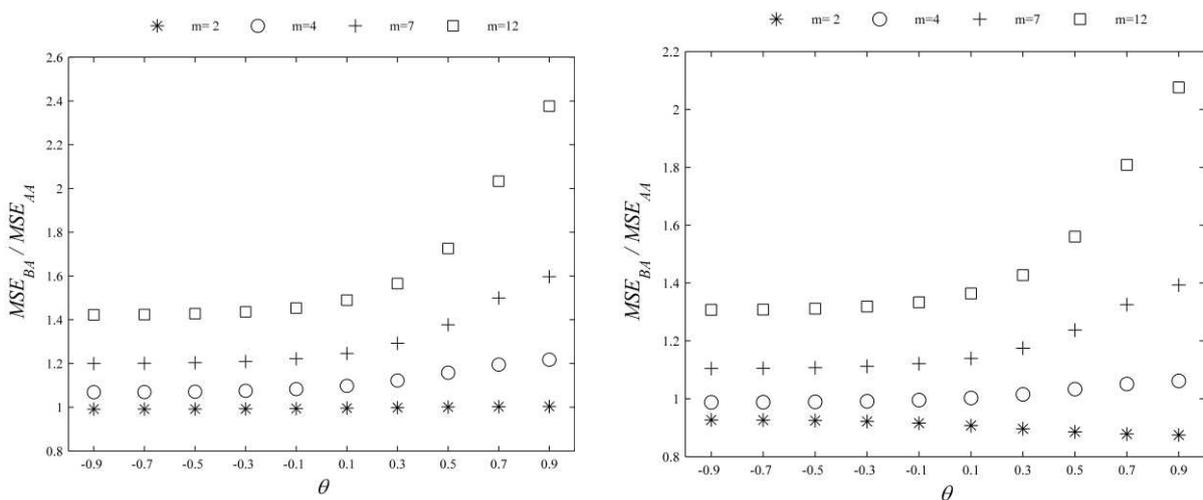
As it is shown in **Figure 3-22a** and **b**, the higher ratio of  $MSE_{BA} / MSE_{AA}$  is associated with higher aggregation level.



**Figure 3-22:** MSE for different values of m, ARIMA(1,0,1) process compared at aggregate level

**3.2.3.2.2 Moving average process order one, ARIMA(0,0,1)**

By considering (3-55) the values of the  $MSE_{BA} / MSE_{AA}$  by varying the aggregation level can be determined. **Figure 3-23a** and **b** show the impact of the aggregation level on the ratio for the whole range of  $\theta$  when  $\alpha=0.1, \beta=0.1$  and  $\alpha=0.1, \beta=0.05$ . It is shown that higher aggregation level is associated with higher ratio.



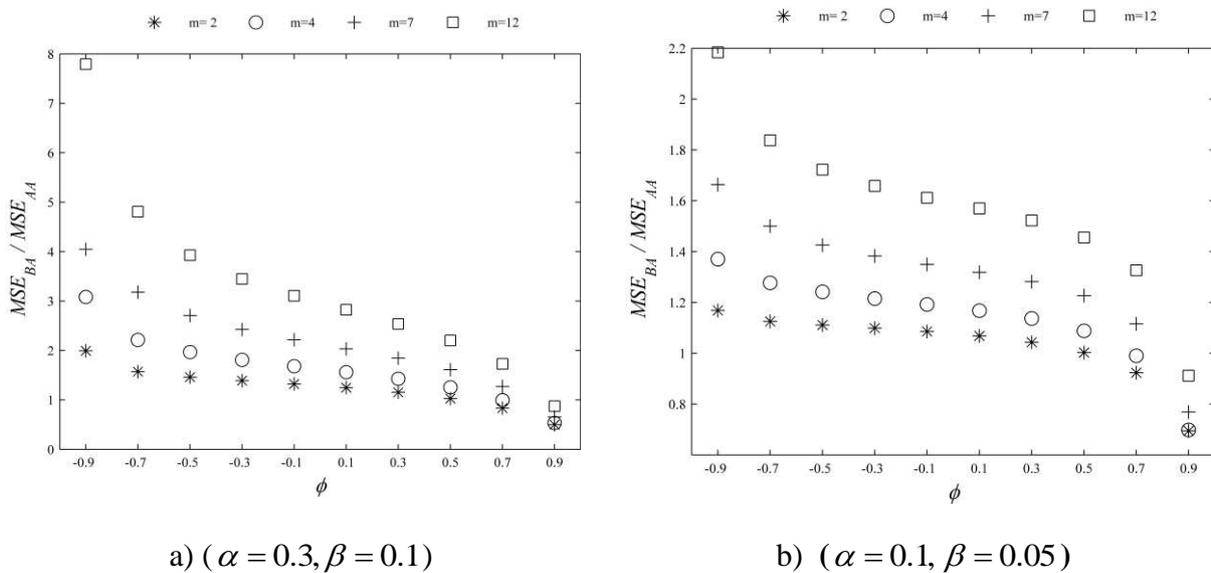
a)  $\alpha=0.1, \beta=0.1$

b)  $\alpha=0.1, \beta=0.05$

**Figure 3-23:** MSE for different values of  $m$ , ARIMA(0,0,1) process comparison at aggregate level

### 3.2.3.2.3 Autoregressive process order one, ARIMA(1,0,0)

**Figure 3-23a** and **b** present the impact of aggregation level on the ratio for the whole range of  $\phi$  when  $\alpha=0.3, \beta=0.1$  and  $\alpha=0.1, \beta=0.05$ . The results show that a higher value of the ratio is achieved by higher aggregation level.



a) ( $\alpha = 0.3, \beta = 0.1$ )

b) ( $\alpha = 0.1, \beta = 0.05$ )

**Figure 3-24:** MSE for different values of  $m$ , ARIMA(1,0,0) process comparison at aggregate level

## 3. Simulation investigation

In this sub-section a simulation experiment based on the theoretically generated data is considered. In this part of the work, simulation analysis is used to test and validate the theoretical results discussed in section 2.

### 3.3.1 Simulation design

Different autoregressive moving average, ARMA type processes are to test the mathematical findings. an ARIMA(1,0,0) process, an ARIMA(0,0,1) process and a mixed ARIMA(1,0,1) process are considered. These processes are analysed in section 1. The disaggregate demands are generated randomly in each period subject to the parameters described in **Table 3-4**. The value of  $\sigma$  is set quite smaller than  $\mu$  to avoid the generation of negative sub-aggregate values. To generate the demands in each period  $t$  that follow ARIMA(1,0,1), ARIMA(0,0,1) and ARIMA(1,0,0), the error terms  $\varepsilon_t$  are first generated randomly. The simulation experiment is designed and run in Matlab 7.10.0. For each parameter combination described in **Table 3-4** a demand series of 1000 observations is generated and 100 replications are introduced.

**Table 3-4:** Parameters of the simulation experiment

$\mu$	$\sigma$	$\alpha, \beta$	$\theta$	$\phi$	N° Replications	N° Time Periods
400	20	0.01: 0.99	-0.9 : +0.9	-0.9 : +0.9	100	1000

The generated series is divided into two parts. The first part (within sample) consists of 450 time periods and is used in order to initialize the *SES* estimates. The second part consists of 550 time periods and is used for the evaluation of the performance (out-of-sample). The values of the smoothing constants before and after aggregation ( $\alpha, \beta$ ) is varied from 0.05 to 0.95 with a step increase of 0.05. For non-aggregation approach, the SES is applied directly to get 550 one-step ahead forecasts and then the variance of the forecast error. is calculated. In order to obtain the forecasts generated by the aggregation approach, first the non-overlapping buckets of aggregate data are created based on a specified aggregation level and then SES method is applied to these aggregate data to get the aggregate forecast. the procedure is explained for the aggregation level equals to two, for higher aggregation level the same procedure is followed. The calculation is begun from the 450<sup>nd</sup> observation in the initial (within sample) part, the observations are summed backwards in buckets of two (2), resulting in an aggregate series consisting of 225 aggregate observations. The average of the aggregate series is obtained and is used as the SES's forecast for the first bucketed period 1. SES is then

applied all the way up to producing a forecast for bucket 226 which gives a forecast for periods 451 and 452. Then the buckets of 2 periods from period 451 backwards are created. Thus, another 225 buckets are created and the very first observation (period 1 in the original data) is not used anymore. The average of these buckets is calculated, it is used as the SES's forecast for the first bucket, then the SES method is applied until the point that a forecast for bucket 227 (periods 452 and 453) is produced. In the next period, the buckets are created backwards from period 452 ending up with 226 buckets and continue like this until obtain the forecasts for 550 periods ahead. Now, if the forecast at the disaggregate level is needed the aggregate forecasts is divided by the aggregation level to get the disaggregate forecast resulted from the aggregate data.

Finally, the value of the variance of the forecast error before aggregation is divided by the variance of the forecast error after aggregation, to obtain the ratio of  $MSE_{BA}$  to  $MSE_{AA}$ .

verification is the process to make sure that no programming error has been made (Kleijnen and Groenendaal, 1992). This can be tested by calculating intermediate results manually and comparing them with the results obtained by the program. This is called tracing (Kleijnen and Groenendaal, 1992). Eyeballing or reading through the code and looking for bugs is another way of verification (Kleijnen and Groenendaal, 1992). The following steps are conducted to verify the simulation model:

- The MATLAB codes are read through to make sure that the correct logic and functions have been used.
- The intermediate and also the final results are compared for a limited number of replications (e.g. 10 replications) with MS Excel.

### 3.3.2 Simulation Result

The simulation results are presented in this sub-section. As discussed in chapter 1, the objective of the simulation analysis I temporal aggregation is to test and validate the results of mathematical evaluation. In section 2 the conditions under which aggregation and non-aggregation approaches may outperform each other are discussed. In the following the results

of simulation analysis are presented for these conditions to compare them with mathematical analysis.

Although a simulation investigation is conducted for all scenarios discussed in the theoretical part but the results of the more general case, ARMA(1,0,1) process are only presented, which has the characteristics of both ARIMA(1,0,0) and ARIMA(0,0,1) process.

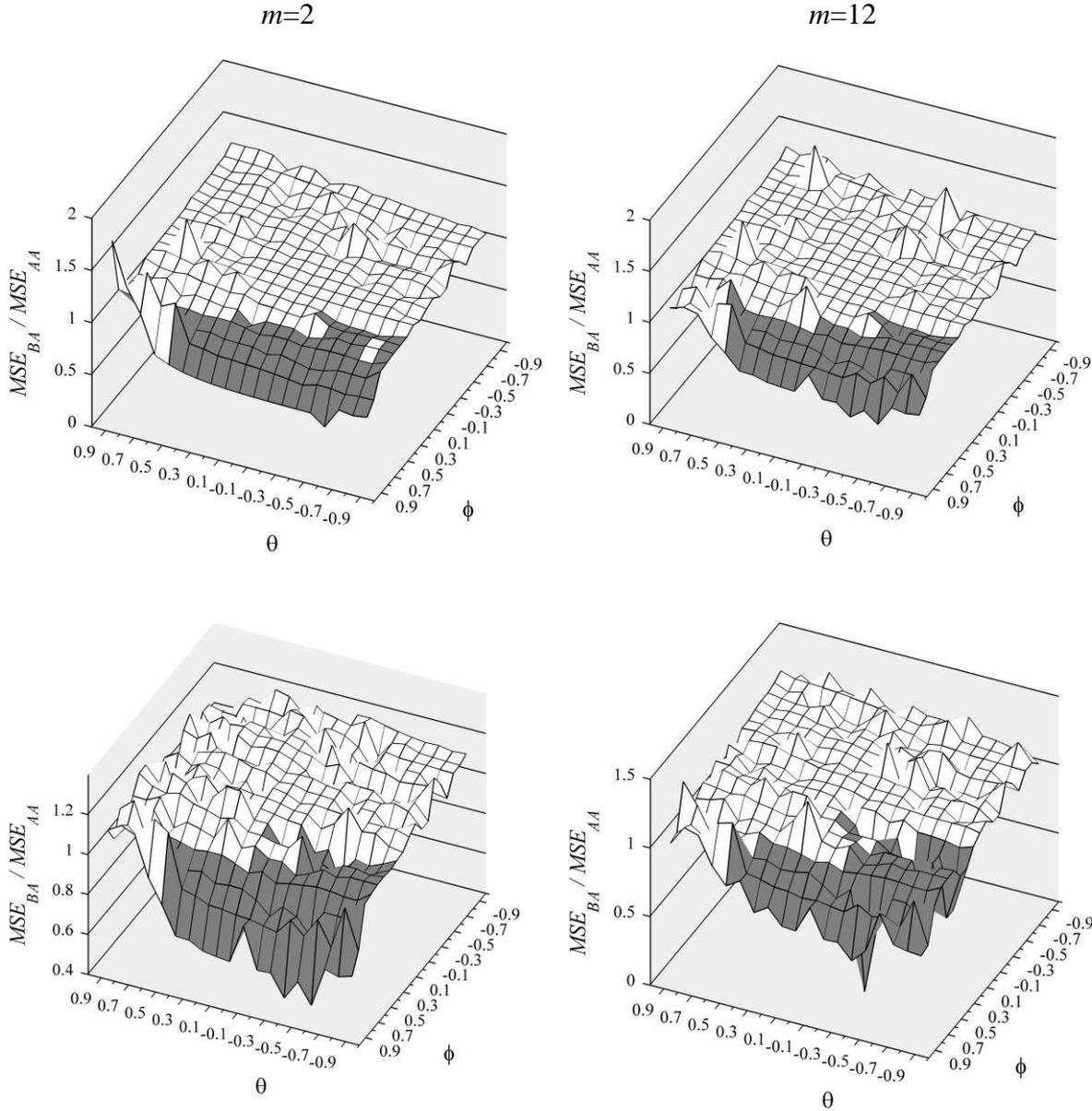
Figure 3-25 presents the impact of the parameters on the ratio of  $MSE_{BA}/MSE_{AA}$  when comparing at disaggregate level for  $m = 2, 12$ ,  $\alpha = 0.1$ , and  $\beta = 0.01, 0.05$ . Shaded areas represent a behavior in favor of the non-aggregation approach.

The simulation results shows that for positive values of  $\phi$  and negative values of  $\theta$ , non-aggregation approach produce more accurate results compared to aggregation approach, however the aggregation approach can provide more accuracy forecasts when  $\phi$  is negative and  $\theta$  is positive. In addition, it is seen that increasing the aggregation level improve the forecasting accuracy when the aggregation approach outperforms the non-aggregation one.

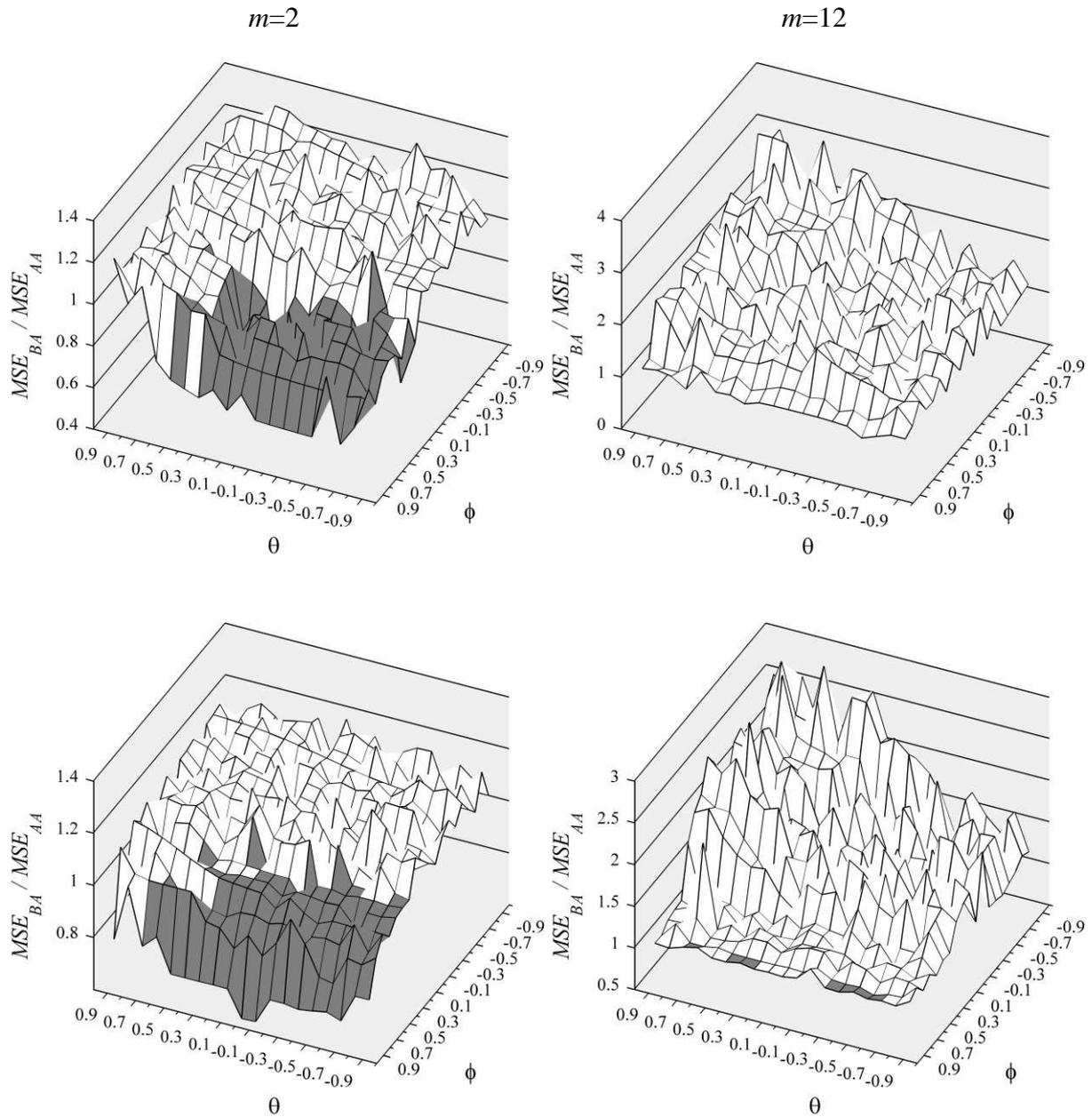
However, by increasing the smoothing constant after aggregation the performance of the aggregation approach deteriorates.

In Figure 3-26 the results of simulation analysis for comparison at the aggregate level are presented for the same parameters used in the previous case. Shaded areas represent a behavior in favor of the non-aggregation approach. As can be observed in Figure 3-26, there is less benefits for the aggregation approach when  $\theta$  takes negative values and  $\phi$  has positive values, and it is seen that for lower aggregation level values, the non-aggregation approach outperforms the aggregation one. However, for higher values of the aggregation level, the aggregation approach outperforms the non-aggregation one regardless of the values of the

autoregressive and the moving average parameters. In addition, by increasing the aggregation level the accuracy of the aggregation approach improves.



**Figure 3-25:** Impact of  $m, \theta, \phi, \alpha$  and  $\beta$  on the ratio of MSE compared at disaggregate level:  $\alpha = 0.1, \beta = 0.01$ (top) ,  $\alpha = 0.1, \beta = 0.05$ (bottom)



**Figure 3-26:** Impact of  $m, \theta, \phi, \alpha$  and  $\beta$  on the ratio of MSE compared at aggregate level:  
 $\alpha = 0.1, \beta = 0.01$  (top) ,  $\alpha = 0.1, \beta = 0.05$  (bottom)

The simulation results presented in both Figure 3-25 and Figure 3-26 generally confirm the results of the theoretical analysis when the underlying series follow an ARIMA(1,0,1) demand process at both disaggregate and aggregate level of comparison.

## 4. Empirical analysis

In this section the empirical validity of the main theoretical findings of this research are assessed. In the following sub-section the details of the empirical data available for the purposes of the investigation along with the experimental structure employed in this work are provided. In sub-section 3.4.2 the actual empirical results are presented.

### 3.4.1 Empirical Dataset and Experiment Details

The demand dataset available for the purposes of this research consists of weekly sales data over a period of two years for 1,798 SKUs from a European grocery store. The Forecast package in R is used to identify the underlying ARIMA demand process for each series and estimate the relevant parameters. It is found that more than 48% of the series may be represented by the processes considered in our research. In particular, 30.26% of the series (544 series) is found to be ARIMA(1,0,0), 12.96% (233 series) to be ARIMA(0,0,1) and 5.06% (91 series) to be ARIMA(1,0,1), (Other popular processes identified are: ARIMA(0,0,0) (16.3%) and ARIMA(0,1,1) (23.7%). This analysis provides some empirical justification on the frequency of stationary, and in particular ARIMA(0,0,1) and ARIMA(1,0,0) processes in real world practices.

In **Table 3-5** and **3-6** and **3-7** the characteristics of the SKUs relevant to this study are summarized by indicating the estimated parameters for ARIMA(1,0,1), ARIMA(0,0,1) and ARIMA(1,0,0) processes. To facilitate a clear presentation, the estimated parameters are grouped in intervals and the corresponding number of SKUs is given for each such interval.

The average  $\theta$  and  $\phi$  value per interval is also presented for the processes respectively. This categorization allows comparing the empirical results with the theoretical findings. It should be noted that the  $\theta$  parameter values are all but one negative and the  $\phi$  parameter values are all but one positive for the ARIMA(0,0,1) and the ARIMA(1,0,0) processes respectively. For the ARIMA(1,0,1) process, the  $\theta$  parameter values are positive or negative and all  $\phi$  parameters are positive, but whole parameters lead to a positive autocorrelation. As such, the data do not cover the entire theoretically feasible range of the parameters. Some studies (Erkip et al., 1990; Lee et al., 1997b; Lee et al., 2000) that have considered empirical

ARIMA(1,0,0) processes, have reported that it is common to have positive correlation/high value of autoregressive parameters in the consumer product industries which is also the case in the dataset used in our research. Replication of our findings in bigger datasets is certainly an avenue for further research.

**Table 3-5:** Processes present in the empirical data set, ARIMA(1,0,1) process

$\theta$ intervals	$\phi$ intervals	Average of $\theta$	Average of $\phi$	Average lag 1 Autocorrelation	No. of SKUs
[0.1,0.5[	[0.6,1[	0.356	0.771	0.5211	23
[0.5,0.9[	[0.6,1[	0.605	0.838	0.3260	39
[-0.2,-0.5[	[0.1,0.5[	-0.328	0.347	0.5631	29
Total number of SKUs:					91

**Table 3-6:** Processes Present in the Empirical Data Set, ARIMA(0,0,1) process

$\theta$ intervals	Average of $\theta$	No. of SKUs
[-.8,-.7[	-0.7252	1
[-.7,-.6[	-0.6329	9
[-.6,-.5[	-0.5393	39
[-.5,-.4[	-0.4471	72
[-.4,-.3[	-0.3509	57
[-.3,-.2[	-0.2520	48
[-.2,-.1[	-0.1989	6
[0,1[	0.2831	1
Total number of SKUs:		233

**Table 3-7:** Processes Present in the Empirical Data Set, ARIMA(1,0,0) process

b) ARIMA(1,0,0)		
$\phi$ intervals	Average of $\phi$	No. of SKUs
[-.1,0[	-0.2240	1
[.1,.2[	0.1981	2
[.2,.3[	0.2534	84
[.3,.4[	0.3549	125
[.4,.5[	0.4479	127
[.5,.6[	0.5512	121
[.6,.7[	0.6433	63
[.7,.8[	0.7352	18
[.8,.9[	0.8256	3
Total number of SKUs:		544

The data series is divided into two parts. The first part (within sample) consists of 62 time periods and is used in order to initialize the *SES* estimates. The second part consists of the remaining 41 time periods and is used for the evaluation of the performance (out-of-sample).

The values of the smoothing constants are varied from 0.05 to 0.95 with a step increase of 0.05. In the classical (non-aggregate) approach, first the 41 one-step ahead forecasts are calculated for each series and then the variance of the forecast error is calculated.

to obtain the forecasts via the aggregation approach, firstly the non-overlapping buckets of aggregate data are created based on a specified aggregation level and then the *SES* method is applied to these aggregate data.

*Aggregation level = 2:* Starting from the 62<sup>nd</sup> weekly observation in the initial (within sample) part, the observations are summed backwards in buckets of two (2), resulting in a bi-weekly series consisting of 31 aggregate observations. The average of aggregate series is

obtained and it is used as the SES's forecast for the first bucket period 1. SES is then applied all the way up to producing a forecast for bucket 32 which is then divided by 2 (the aggregation level,  $m=2$ ) and it gives a forecast for periods 63 and 64. The forecast for period 64 is dropped and those of 63 is recorded (they are equal anyway). Then the buckets of 2 periods from period 63 backwards are created. Therefore, another 31 buckets are created and the very first observation (period 1 in the original data) is not used anymore. The average of these buckets are calculated (they are different from those created before), that average is used as the SES's forecast for the first bucket, the forecasting process is continues using SES until the point that a forecast for bucket 32 (periods 64 and 65) is obtained. The forecast for period 64 is kept and so on. In the next period, the buckets are created backwards from period 64 ending up with 32 buckets and continue like this until obtain the forecasts for 41 periods ahead.

Aggregation level = 3 . . . 24: Similarly, the same procedure is followed with time buckets of up to 24 periods. At this point there are 2 aggregate biweekly observations ( $2 \times 24 = 48$ ), thus 14 weekly observations at the start of the original series remain unused.

Finally, the value of the variance of the forecast error before aggregation is divided by the variance of the forecast error after aggregation to obtain the ratio of  $MSE_{BA}$  to  $MSE_{AA}$ .

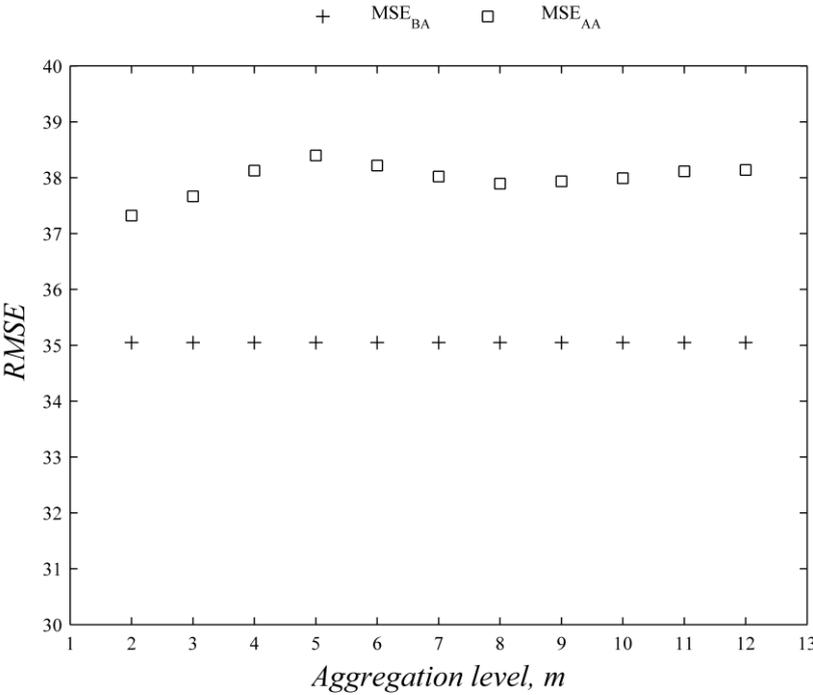
### 3.4.2 Empirical Results

In this section the results of empirical investigation compared at both disaggregate and aggregate level for all processes under consideration are presented.

#### 3.4.2.1 Comparison at disaggregate level

In the first part, the validity of mathematical results are evaluated by real data set when the non-aggregate demand follow an ARIMA(1,0,1) process and the comparison is conducted in the disaggregate level. The empirical results show that when the optimal smoothing constant values  $\alpha$  and  $\beta$  are used, then for all values of aggregation level  $m$  the non-aggregation approach outperforms the aggregation one. This is in agreement with our findings as the real data set presented in Table 3-5 takes positive autocorrelation, not only for lag1 but

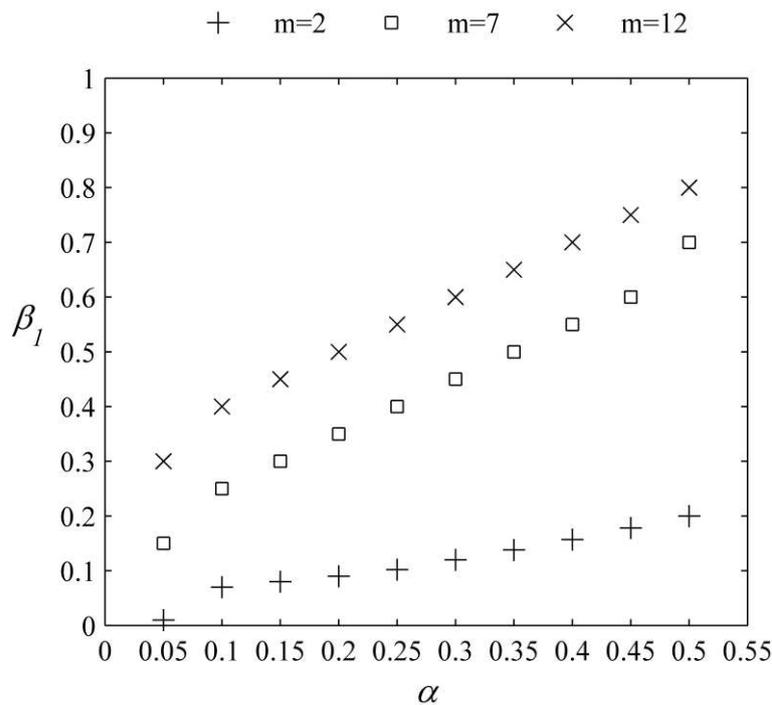
it spans over longer lags. According to the theoretical findings when the autocorrelation is positive the non-aggregation approach performs better and no level of aggregation improve the performance of the aggregation approach. As it is shown in **Figure 3-27**, for all values of  $m$  the  $MSE_{BA}$  is lower than  $MSE_{AA}$ . It should be noted that the results is presented based on the  $RMSE$ (root mean square error) which is similar to  $MSE$ . The  $MSE$  reduction can be as high as 8% for the aggregation approach.



**Figure 3-27:** Empirical results compared at disaggregate level, ARIMA(1,0,1) process

In sub-sections 3.2.2.1.2 and 3 the conditions under which the aggregate forecasts may perform better than the non-aggregate are analytically examined by the ratio of  $MSE_{BA}$  to  $MSE_{AA}$ . The cut-off points of the smoothing constant of the aggregate series  $\beta$  that should be used (i.e. any value of  $\beta$  that is lower than the cut-off point  $\beta_1$  implies an outperformance of the aggregation approach) have also been determined for both the ARIMA(0,0,1) and ARIMA(1,0,0) process. In the following figures the results of the empirical analysis for these processes are presented. Additionally, the degree to which they validate the theoretical findings is investigated.

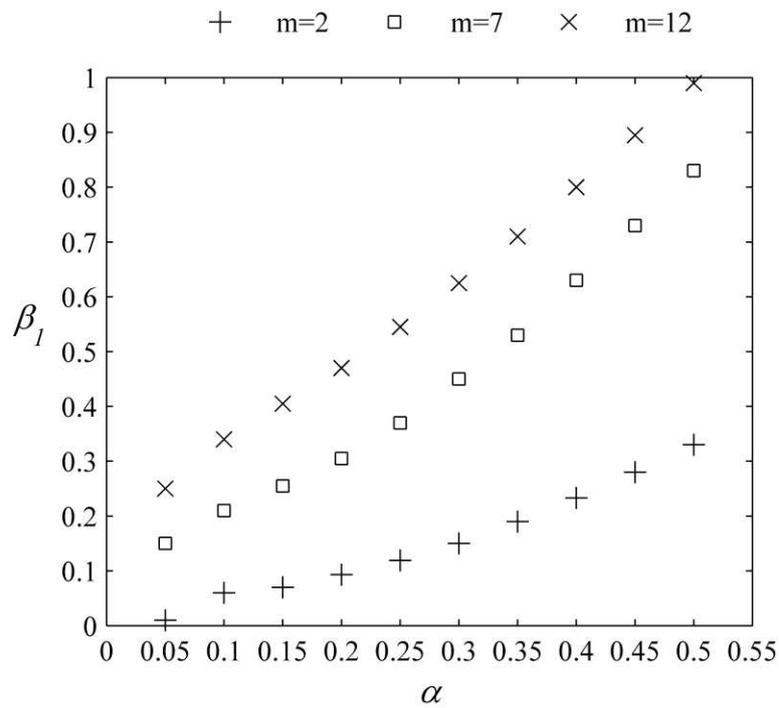
In **Figure 3-28**, the cut-off point  $\beta_1$  is presented for a fixed values of  $\alpha$  and  $m$  when the non-aggregate demand of the SKUs follows an ARIMA(0,0,1) process. Please recall that the cut-off point  $\beta_1$  is the value below which any  $\beta$  value implies that the aggregation approach outperforms the non-aggregation one. Note that the results for  $\alpha \leq 0.5$  are only presented since this range is viewed as realistic for the stationary processes considered in this work.



**Figure 3-28:** Cut-off points of  $\beta$  implying an outperformance of the aggregation approach for different values of  $\alpha$  and  $m$  compared at disaggregate level, ARIMA(0,0,1) process

The empirical results show that for a low aggregation level  $m=2$ , the cut-off point is relatively low since  $\beta_1=0.2$  for a relatively high  $\alpha$  value equal to 0.5. In that case, the MSE reduction when  $\beta=0.05$  is equal to 8.89% and the MSE ratio decreases for higher values of  $\beta$ . Obviously, the cut-off value considerably increases when the aggregation level increases. For example, when we consider the aggregation level  $m=12$ , the cut-off point may go up to  $\beta_1=0.8$  for  $\alpha$  value equal to 0.5. In that case the MSE reduction when  $\beta=0.05$  is equal to 12.13%. This shows the considerable region where the aggregation approach outperforms the non-aggregation one for high aggregation levels. Hence, increasing the aggregation level

improves the performance of the aggregation approach and the best results can be achieved for small values of  $\beta$  and high aggregation levels  $m$ . These empirical results generally confirm the theoretical findings.



**Figure 3-29:** Cut-off points of  $\beta$  implying an outperformance of the aggregation approach for different values of  $\alpha$  and  $m$  compared at disaggregate level, ARIMA(1,0,0) process with  $-1 < \phi \leq 0.33$ .

Figure 3-29 shows the cut-off point  $\beta_1$  for fixed values of  $\alpha$  and  $m$  when the SKUs have a non-aggregate demand that follows an ARIMA(1,0,0) process with  $-1 < \phi \leq 0.33$ . The empirical results show that for a low aggregation level  $m=2$ , low  $\beta$  values should be selected in order to have an outperformance of the aggregation approach. For example when an aggregation level  $m=2$  is used, the cut-off point  $\beta_1=0.33$  for an  $\alpha$  value equal to 0.5 and the MSE reduction when  $\beta=0.05$  is equal to 12.45%. The cut-off points considerably increase when the aggregation level increases. Figure 3-29 shows also that for an  $\alpha$  value equal to 0.5 and when the aggregation level  $m=12$ , the cut-off point  $\beta_1$  is almost equal to 1, which means that the aggregation approach always outperforms the non-aggregation one in that case. That

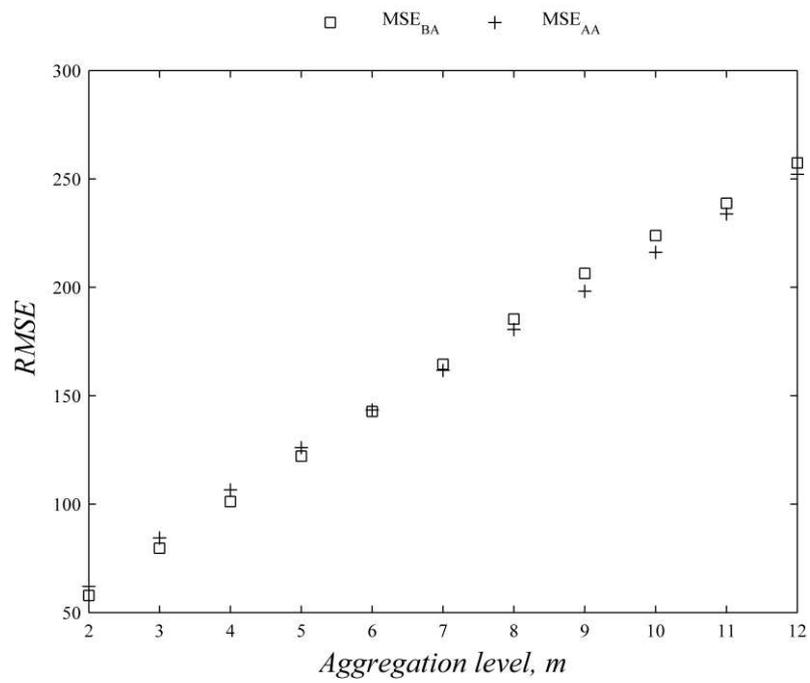
results also in a MSE reduction equal to 15.11% that decreases for higher values of  $\beta$ . However, it should be noted that for the SKUs where  $0.33 < \phi < 1$ , the empirical results show that when the optimal value of  $\alpha$  is used for all values of  $\beta$  and  $m$ , the non-aggregation approach outperforms the aggregation one.

The empirical analysis confirms overall the results of the theoretical evaluation both for all processes under consideration. What can be concluded here is that there is a considerable range of the values of the smoothing constant of the aggregate series that implies a benefit of using the aggregation approach. This benefit can also be substantial for high aggregation levels and low smoothing constants. Note that such analysis can be utilized as an indicator on when the aggregation approach should be used and which parameters lead to the outperformance of this approach.

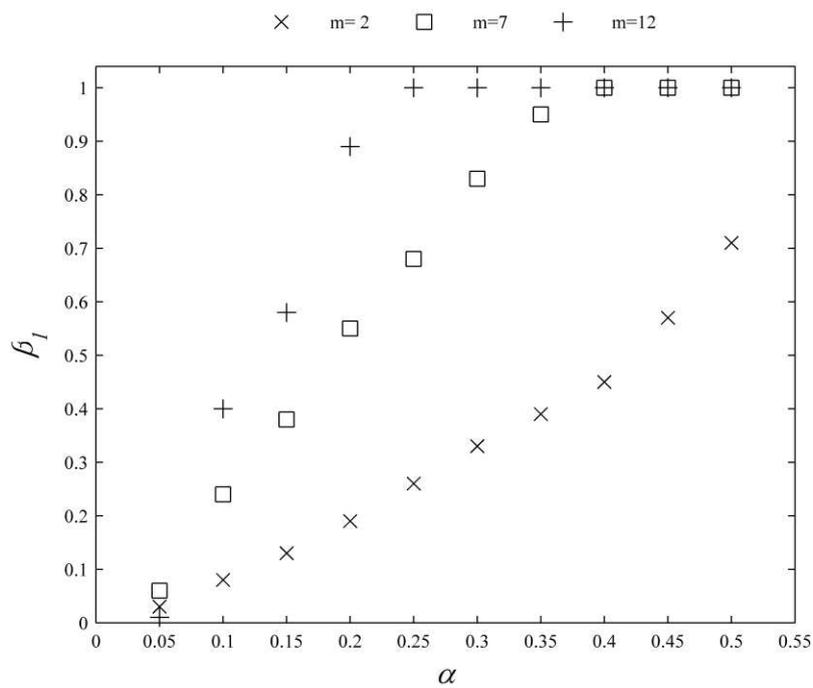
### 3.4.2.2 Comparison at aggregate level

In this part the validity of the findings in forecasting the aggregate demand is tested by real data sets. In sub-sections 3.2.1.2 and 3.2.2.2 the superiority conditions of the aggregation and non-aggregation approaches are identify when a cumulative  $m$  step ahead forecast is required. It is shown that for positive autocorrelation associated with low aggregation level, non-aggregation approach works better but by increasing the aggregation level the performance is improved even for high positive autocorrelation.

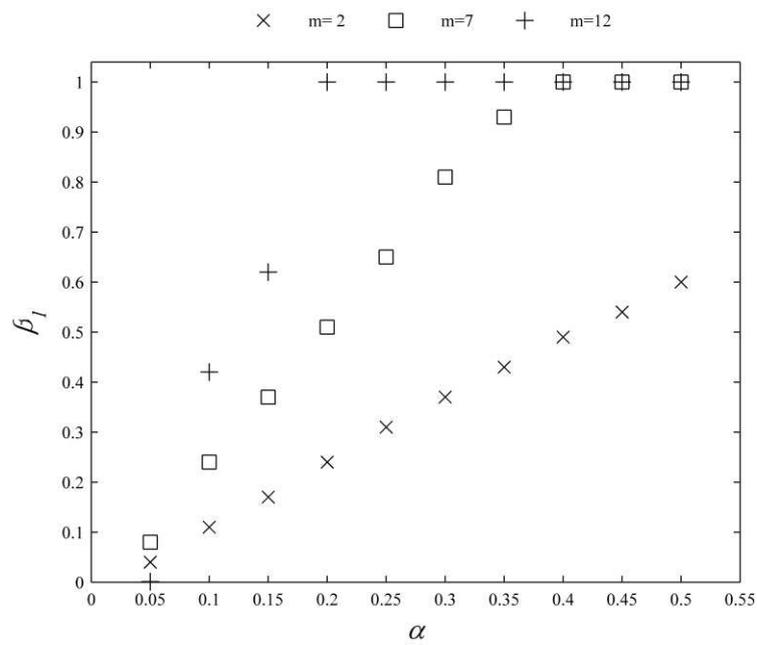
**Figure 3-30** shows the results of both aggregation and non-aggregation approaches for different values of aggregation level  $m$  when the optimal smoothing constants before and after aggregation is used and the non-aggregate demand series follow an ARIMA(1,0,1) process. The results show that for the aggregation level  $m$  up to six, the  $MSE_{BA}$  is smaller than  $MSE_{AA}$ . However, as  $m$  takes higher values than six, the latter becomes smaller. Therefore, the empirical results show that when the non-aggregate demand follow an ARIMA(1,0,1) process and autocorrelation is positive (refer to Table 3-5) then for lower values of  $m(m \leq 6)$  the non-aggregation approach works better. However for  $m > 6$ , aggregation approach outperforms the non-aggregation one.



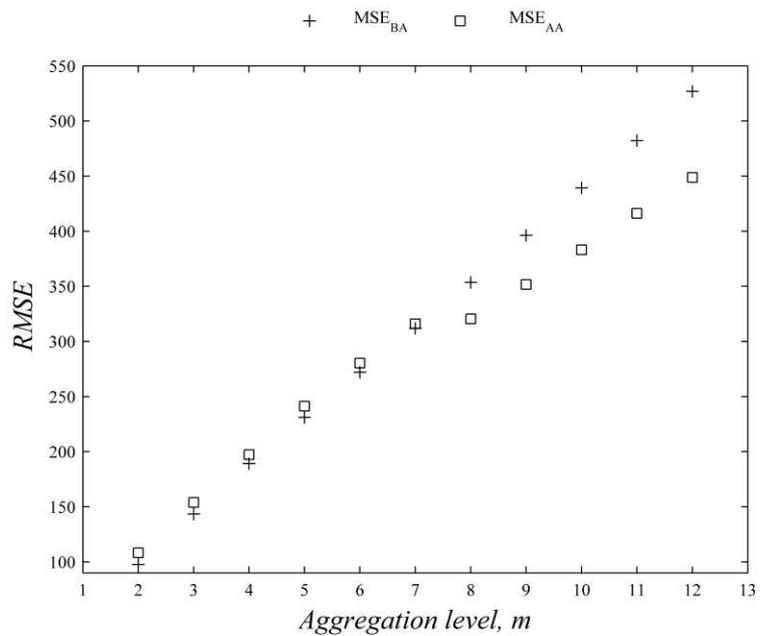
**Figure 3-30:** Empirical results compared at aggregate level, ARIMA(1,0,1) process



**Figure 3-31:** Cut-off points of  $\beta$  implying an outperformance of the aggregation approach for different values of  $\alpha$  and m compared at aggregate level, ARIMA(0,0,1) process.



**Figure 3-32:** Cut-off points of  $\beta$  implying an outperformance of the aggregation approach for different values of  $\alpha$  and  $m$  compared at aggregate level, ARIMA(1,0,0) process with  $-1 < \phi < 0.33$ .



**Figure 3-33:** Cut-off points of  $\beta$  implying an outperformance of the aggregation approach for different values of  $\alpha$  and  $m$  compared at aggregate level, ARIMA(1,0,0) process with  $0.33 < \phi < 1$ .

**Figure 3-31** presents the results of empirical analysis compared at aggregate level when the non-aggregate demand follows an ARIMA(0,0,1) process. It is shown that there is always a value of smoothing constant after aggregation  $\beta$  for which the aggregation approach outperforms non-aggregation one. The cut-off point  $\beta_1$  is determined for fixed values of  $\alpha$  and  $m$  when the non-aggregate demand of the SKUs follows an ARIMA(0,0,1) process. The results show that for high values of aggregation level, the aggregation approach outperforms the non-aggregation one for a wide range of  $\beta$  values.

**Figure 3-32** shows the cut-off point  $\beta_1$  for fixed values of  $\alpha$  and  $m$  when the non-aggregate SKUs follow an ARIMA(1,0,0) process with  $-1 < \phi \leq 0.33$ . As it is discussed above, for these autoregressive values the autocorrelation is not highly positive. For these range of  $\phi$  there is always a value of  $\beta$  for which the aggregation approach outperforms the non-aggregation one. The results show that the cut-off points considerably increase when the aggregation level increases. **Figure 3-32** shows also that for an  $\alpha$  value greater than and equal to 0.2 and when the aggregation level  $m=12$ , the cut-off point  $\beta_1$  is almost equal to 1, which means that for these values aggregation approach always performs better. The MSE reduction associated with these values and the smoothing constant after aggregation  $\beta=0.01$  can be as high as 50%.

In **Figure 3-33**, the forecast results of SKUs with  $0.33 < \phi < 1$  are presented. The empirical results show that when the optimal value of  $\alpha$  and  $\beta$  is used, then for lower value of  $m$ , the  $MSE_{BA}$  is smaller than the  $MSE_{AA}$ . However, as the aggregation level increases the latter becomes smaller than the former one and consequently the non-aggregation approach outperforms the aggregation one. The results show that for the aggregation level  $m$  smaller than or equal to seven the non-aggregation approach performs better, but for the values of  $m$  greater than seven, the aggregation approach outperforms the non-aggregation one. These results confirm the results of analytical evaluation presented in sub-sections 3.2.1.2.3 and 3.2.2.2.3.

## 5. Conclusion

In this chapter the impact of temporal aggregation on demand forecasting has been evaluated by theoretical, simulation and empirical investigation. The evaluations were based on the consideration of the Mean Squared Error (MSE) before and after aggregation ( $MSE_{BA} / MSE_{AA}$ ) and comparisons were undertaken at both disaggregate and aggregate demand level. It is assumed that the demand follow an ARIMA(1,0,1), ARIMA(0,0,1) and ARIMA(1,0,0) process and a Single Exponential Smoothing is used as a forecasting method. The conditions under which the aggregation approach outperforms the non-aggregation are identified.

The results show that when the autocorrelation is highly positive in the original series the non-aggregation approach may outperform the aggregation one. In general there are fewer benefits for the aggregation approach with high positive autocorrelation than the series with low positive or negative autocorrelation. This is an intuitive finding since when the autocorrelation is highly positive, at any time the most recent demand information is vital. Therefore, in that case the disaggregate process works better as it fully exploits such recent information. However, on the contrary, for low positive autocorrelation or negative autocorrelation, the recent demand information is not that crucial then a more long term view on demand is preferable. As discussed above it can be obtained by selecting high aggregation levels and low smoothing constants. This is also an important empirical insight since managers may know what to expect (in terms of any potential gains) based on the autocorrelation levels present in their series.

When the demand process follow either an ARIMA(1,0,1) or an ARIMA(1,0,0) process associated with high positive autocorrelation, and the comparison is undertaken at disaggregate level, the results show that non level of aggregation improve the accuracy so the non-aggregation approach always outperforms the aggregation one. However, when comparing is undertaken at aggregate level, for low aggregation level the non-aggregation approach may outperforms the aggregation one, but for higher values of aggregation level, the aggregation approach always provide more accurate forecasts.

It is also found that the performance of the aggregation approach improves as the smoothing constant value employed at the aggregate series reduces and the aggregation level increases. This is true for both comparison at disaggregate and aggregate level.

## Chapter 4 Cross-Sectional Aggregation

---

In chapter 3, the effect of the non-overlapping temporal aggregation on demand forecasting is analysed. In this chapter the effectiveness of cross-sectional approach on demand forecasting is evaluated. An important decision involved in the forecasting process is the determination of the degree of aggregation that forecasts should refer to with respect to the number of products involved. The hierarchical level at which forecasting is performed it depends on the function the forecasts are fed into. There are several examples with regards to products (or Stock Keeping Units - SKUs) in particular: *i*) forecasting at the individual SKU level is required for inventory control, *ii*) product family forecasts may be required for Master Production Scheduling, *iii*) forecasts across a group of items ordered from the same supplier may be required for the purpose of consolidating orders, and *iiii*) forecasts across the items sold to a specific large customer may determine transportation and routing decisions etc.

One intuitively appealing approach to obtain higher level forecasts is by *cross-sectional (also referred to as hierarchical) aggregation*, which involves aggregating different items (i.e. aggregating the requirements for different items usually in one specific period) to reduce variability. Existing approaches to the cross-sectional forecasting usually involve either a bottom-up (BU) or a top-down (TD) approach (or a combination of the two). When forecasting at the aggregate level is of interest, the former involves the aggregation of individual SKU forecasts to the group level whereas the latter relates to forecasting directly at the group level (i.e. first aggregate requirements and then extrapolate directly at the aggregate level). When the emphasis is on forecasting at the subaggregate level, then the BU approach relates to direct extrapolation at the subaggregate level whereas the TD involves the disaggregation of the forecasts produced directly at the group level. An important issue that has attracted the attention of many researchers as well as practitioners over the last few decades is the effectiveness of such cross-sectional forecasting approaches.

TD and BU forecasting approaches are extremely useful towards improving the accuracy of forecasts and plans when leveraged within an S&OP (Sales and Operations Planning) process (Lapide, 2006). The S&OP is a multi-functional process that involves managers from all departments (Sales, Customer Service, Supply Chain, Marketing, Manufacturing, Logistic, Procurement and Finance), where each department requires different levels of demand forecasts (Lapide, 2004). For example, in marketing, forecasting of revenues by product groups and brands is needed, sales departments deal with sales forecasts by customer accounts and/or sales channels. Supply chain managers request SKU level forecasts, while finance department requires forecasts that are aggregate into budgetary units in terms of revenues and costs (Bozos and Nikolopoulos, 2011).

In this chapter, the relative effectiveness of the BU and TD approach for forecasting is evaluated. It is recognized that forecasts may be equally required at both the aggregate and subaggregate level, and as such comparisons are performed at both levels. The effectiveness of the two approaches is analytically studied when the underlying series follows either a stationary first order Autoregressive Moving Average process  $ARIMA(1,0,1)$  or a non-stationary Integrated Moving Average process of order one,  $ARIMA(0,1,1)$ , and the forecasting method is the Single Exponential Smoothing (SES) which is the optimal estimator for the  $ARIMA(0,1,1)$  process (Box et al., 2008).

The assumptions bear a significant degree of realism. An ARMA(Autoregressive Moving Average) model often fits demand data better than an autoregressive or moving average model alone, since typically demand contain structurally both moving average and autoregressive characteristics (Duc et al., 2008). The ARMA process have been found to fit demand for long lifecycle goods such as fuel, food products, machine tools, etc (Chopra and Meindl, 2001; Nahmias, 1993). It has also been shown that the ARMA demand processes occur naturally in multi-stage supply chains (Gaur et al., 2005; Zhang, 2004). There is also considerable evidence to suggest that inventory demand is non-stationary and thus relevant processes should be assumed for representing their underlying structure. Tunc et al. (2011) stated that non-stationary stochastic demands are very common in all industrial settings associated with seasonal patterns, trends, business cycles, and limited-life items such as the high-tech industry (Chien et al., 2008; Graves and Willems, 2000, 2008) and grocery distribution (Erkip et al., 1990; Lee et al., 1997a; Martel et al., 1995). There is also some

evidence that demand may follow an ARIMA(0,1,1) process in particular (which is the process considered in this study). This process has often been found to be useful in inventory control problems and econometrics (Box et al., 2008). In addition, Mahajan and Desai (2011) stated that retailers often face a non-stationary demand that follows an ARIMA(0,1,1) process.

Moreover, SES is a very popular forecasting method in industry (Acar and Gardner, 2012; Gardner, 1990, 2006; Taylor, 2003). In terms of the practical relevance of this research we refer to a set of SKUs where a large proportion of them follow an ARIMA(1,0,1) or an ARIMA (0,1,1) processes. This is not an untypical scenario as demonstrated by analysis of empirical datasets including our own empirical investigation.

The mathematical analysis is complemented by a numerical experiment to evaluate in detail the conditions under which one approach outperforms the other. Such an experiment also allows the introduction of non-identical process parameters of the subaggregate series and the comparison at the subaggregate level when the subaggregate items follow an ARIMA(0,1,1) process. In addition, an empirical investigation is also conducted to assess the validity of the results on real data from a European superstore.

Considerable part of this chapter is presented in Rostami-Tabar et al (2013d) and Rostami-Tabar et al(2013b).

The remainder of this chapter is structured as follows. In section 1 the assumptions and notations used in this study are described, In addition an analytical evaluation of the variance of the forecast error related to both the BU and TD approaches is conducted. In section 2 the analytical results are presented. Next a simulation study is conducted following the simulation results in section 3. Finally, an empirical investigation is conducted in Section 4 and the chapter concludes in Section 5.

## **1. Theoretical analysis**

In this section the variance of the forecast error associated with the TD and BU approaches is derived. Comparisons may be performed at both the aggregate and the subaggregate level although in this theoretical analysis for the ARIMA(0,1,1) process, the

comparisons are performed only at the former level since results regarding the latter are intractable. However, in the simulation study following the theoretical analysis, various assumptions are relaxed and the results for the ARIMA(0,1,1) and the ARIMA(1,0,1) processes are presented. The comparison is undertaken at both subaggregate and aggregate level.

When forecasting at the aggregate level is of interest, the former involves the aggregation of individual SKU forecasts to the group level whereas the latter relates to forecasting directly at the group level (i.e. first aggregate requirements and then extrapolate directly at the aggregate level). When the emphasis is on forecasting at the subaggregate level, then the BU relates to direct extrapolation at the subaggregate level whereas the TD involves the disaggregation of the forecasts produced directly at the group level.

#### 4.1.1 Notation and assumptions

For the remainder of the paper, the following notations are used:

$d_{i,t}$ : Subaggregate demand  $i$  in period  $t$

$\rho_{i,j}$ : Correlation between the error term of subaggregate item  $i$  and  $j$  (cross-correlation)

$D_t$ : Aggregate demand in period  $t$

$\varepsilon_{i,t-k}$ : Independent random variable for subaggregate demand  $i$  in period  $t$ , normally distributed with zero mean and variance  $\sigma^2$

$\varepsilon'_{t-k}$ : Independent random variable for aggregate demand in period  $t$ , normally distributed with zero mean and variance  $\sigma'^2$

$f_{i,t}$ : Forecast of subaggregate demand in period  $t$ , the forecast produced in  $t-1$  for the demand in  $t$ .

$F_t$ : Forecast of aggregate demand in period  $t$ , the forecast produce in  $t-1$  for the demand in  $t$ .

$\alpha_i$ : Smoothing constant used in the Single Exponential Smoothing method for each subaggregate demand in the BU approach,  $0 \leq \alpha_i \leq 1$

$\alpha_{TD}$ : Smoothing constant used in the Single Exponential Smoothing method for aggregate demand in TD approach,  $0 \leq \alpha_{TD} \leq 1$

$p_i$ : the relative weight of subaggregate item  $i$ 's contribution to the aggregate family, where  $\sum_{i=1}^N p_i = 1$ .

$V_{BU}$ : Variance of Forecast Error of the BU approach

$V_{TD}$ : Variance of Forecast Error of the TD approach

$\theta_i$ : Moving average parameter of subaggregate demand  $i$ ,  $|\theta_i| < 1$

$\phi_i$ : Autogressive parameter of subaggregate demand  $i$ ,  $|\phi_i| < 1$

$\theta'$ : Moving average parameter of aggregate demand,  $|\theta'| < 1$

$\mu$ : Expected value of sub-aggregate demand  $i$  in any time period

$\mu'$ : Expected value of aggregate demand in any time period

It is assumed that all the subaggregate demand series  $d_{i,t}$  follow either a first order autoregressive moving average, ARIMA(1,0,1), or a first order Integrated Moving Average process, ARIMA(0,1,1). This can be mathematically written in period  $t$  by (1) and (2) respectively:

$$d_{i,t} = t\mu_i + d_{i,t-1} + \varepsilon_{i,t} - \theta_i \varepsilon_{i,t-1} \quad i=1,2,\dots,N \quad (4-1)$$

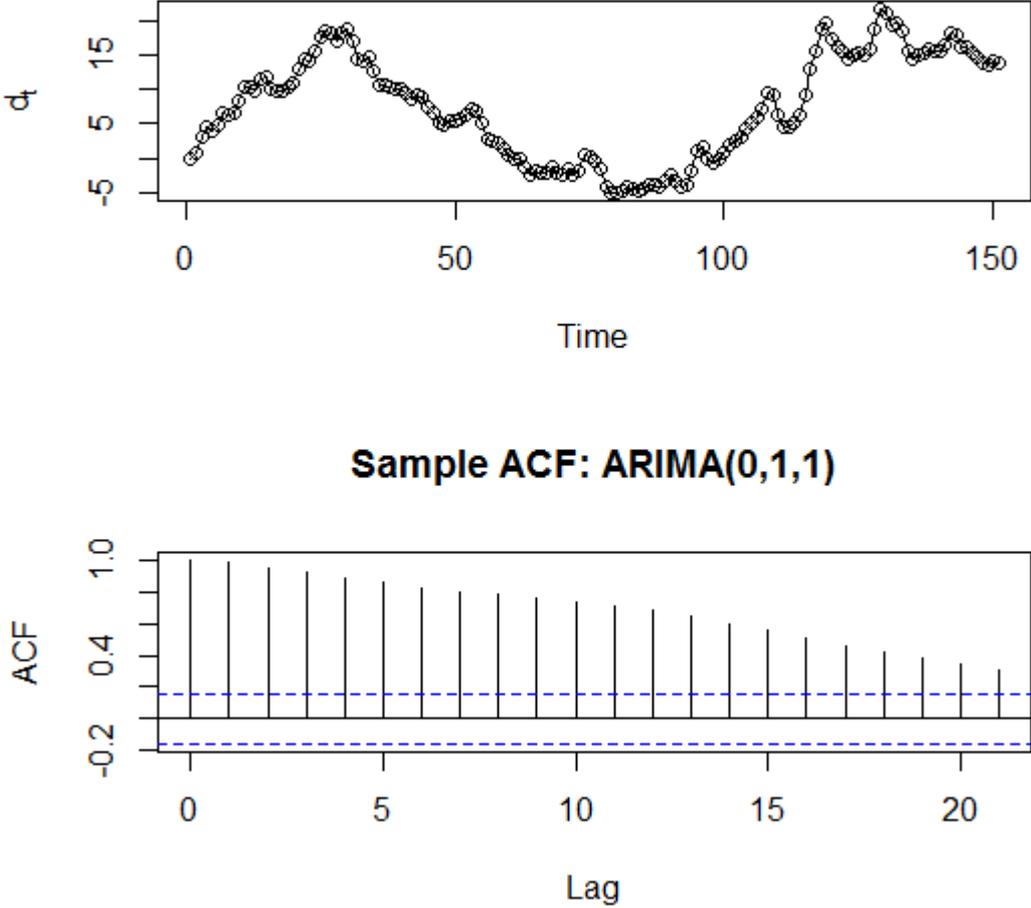
$$d_{i,t} = \mu_i(1 - \phi_i) - \theta_i \varepsilon_{i,t-1} + \phi_i d_{i,t-1} + \varepsilon_t. \quad (4-2)$$

From (1) it is clear that the demand in the next period is the demand in the current period plus an error term. By expanding (4-1) we have:

$$d_{i,t} = t\mu_i + \varepsilon_{i,t} + \alpha_i \varepsilon_{i,t-1} + \alpha_i \varepsilon_{i,t-2} + \dots + \alpha_i \varepsilon_{i,1} \quad (4-3)$$

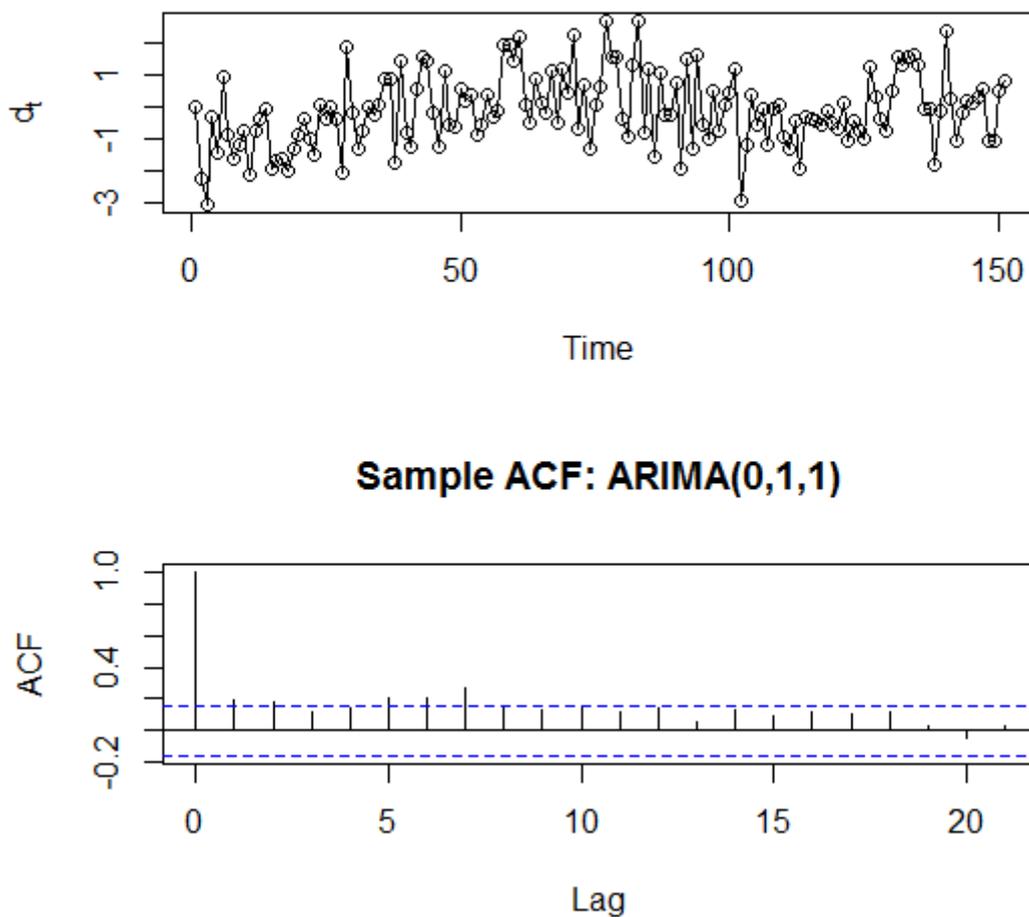
where  $\alpha_i = 1 - \theta_i$ . It should be noted that only under this condition on  $\alpha_i$ , SES is optimal as it provides the minimum mean square forecasts for the ARIMA(0,1,1) process.

Here the smoothing constant values are considered as a control parameter determined by forecasters that varies between zero and one. Obviously, since  $0 \leq \alpha_i \leq 1$ , under this condition  $\theta_i$  (only for ARIMA(0,1,1) process) only takes the values between zero and one and does not cover the whole range of  $-1 \leq \theta_i \leq 1$ . However, the theoretical analysis is still valid for the whole range of  $-1 \leq \theta_i \leq 1$ . In addition, in the simulation analysis this assumption to cover the whole range of  $-1 \leq \theta_i \leq 1$  are relaxed when the value of the smoothing constant is fixed.



**Figure 4-1:** Sample autocorrelation of ARIMA(0,1,1) process when  $\theta = -0.9$ .

When the underlying process follows an ARIMA(0,1,1) process, as  $\theta$  moves from +1 toward -1 the resulting underlying structure changes considerably. When  $0.5 < \theta < -1$ , the autocorrelation is highly positive and it spans all time lags (not only lag 1). For example, for  $\theta = -0.9$  the autocorrelation is very close to +1 with smooth exponential decay by increasing the lags (see **Figure 4-1**). As we move up towards  $\theta \approx +1$  the autocorrelation reduces but still remains positive and for high positive values of  $\theta$  it becomes close to zero meaning that the series are random (see **Figure 4-2**).



**Figure 4-2:** Sample autocorrelation of ARIMA(1,0,1) process when  $\theta = 0.9$ .

However, the behaviour of the ARIMA(1,0,1) process is different with those of the ARIMA(0,1,1) process by changing the parameters. For different combinations of the process parameters, the resulting underlying structure changes considerably.

When the demand follows an ARIMA(1,0,1) process the auto-covariance function is (Box et al., 2008):

$$\gamma_{i,k} = \begin{cases} \frac{(1-2\phi_i\theta_i + \theta_i^2)}{1-\phi_i^2} \sigma_i^2 & k=0 \\ \frac{(\phi_i - \theta_i)(1-\phi_i\theta_i)}{1-\phi_i^2} \sigma_i^2 & |k|=1, \\ \phi_i \gamma_{i,k-1} & |k|>1 \end{cases} \quad (4-4)$$

When the demand follows an ARIMA(1,0,1) process the auto-covariance and autocorrelation functions are (Box et al., 2008). When the demand follows an ARIMA(1,0,1) process the auto-covariance and autocorrelation functions are (Box et al., 2008).

When the demand follows an ARIMA(1,0,1) process the auto-covariance and autocorrelation functions are (Box et al., 2008):

$$\gamma_k = \begin{cases} \frac{(1-2\phi\theta + \theta^2)}{1-\phi^2} \sigma^2 & k=0 \\ \frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \sigma^2 & |k|=1, \\ \phi \gamma_{k-1} & |k|>1 \end{cases} \quad (3-2)$$

$$\rho(k) = \begin{cases} \frac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta + \theta^2} & |k|=1 \\ \phi & |k|>1 \end{cases} \quad (3-3)$$

For different combinations of the process parameters, the resulting underlying structure changes considerably. **Table 3-1** presents the autocorrelation structure for different

process parameters which helps to better understand the process and can be useful to interpret the results of the forthcoming analysis

For different combinations of the process parameters, the resulting underlying structure changes considerably. **Table 3-1** presents the autocorrelation structure for different process parameters which helps to better understand the process and can be useful to interpret the results of the forthcoming analysis.

It is assumed that all the subaggregate demand process parameters are identical ( $\theta_1 = \theta_2 = \theta_3 = \dots = \theta_N$ ). This assumption is considered only for the purpose of the theoretical analysis and, as above, it is also relaxed in the simulation part of this work. The concerned assumption implies that the aggregate demand also follows the same process as subaggregate items. If  $\theta_1 \neq \theta_2 \neq \theta_3 \neq \dots \neq \theta_N$  then the sum of the subaggregate items is not necessarily the same process (Lütkepohl, 1984).

The aggregate demand in period  $t$ ,  $D_t$  can be expressed as the sum of the demands of the subaggregate items, i.e.  $D_t = \sum_{i=1}^N d_{i,t}$ .

The forecasting method considered in this study is the Single Exponential Smoothing (SES). This method is being applied in many companies. Due to its simplicity, It has been specifically applied in an inventory production planning environment (Gardner, 1990). Using SES, the forecast of subaggregate demand  $i$  in period  $t$  produced at the end of period  $t-1$  is

$$f_{i,t} = \sum_{k=1}^{\infty} \alpha_i (1 - \alpha_i)^{k-1} d_{i,t-k}, \quad . \quad (4-5)$$

The forecast of subaggregate item  $i$  in period  $t$  for the ARIMA(0,1,1) process can be expressed as a function of the error terms as follows:

$$f_{i,t} = t\mu_i + \alpha_i \varepsilon_{i,t-1} + \alpha_i \varepsilon_{i,t-2} + \dots + \alpha_i \varepsilon_{i,1} \quad (4-6)$$

It is further assumed that the standard deviation of the error term in (4-1) and (3-7) above is significantly smaller than the expected value of the demand. Thus, when demand is generated, the probability of a negative value is negligible.

#### 4.1.2 Variance of forecast error at aggregate level

The variance of forecast error corresponding to the TD ( $V_{TD}$ ) and the BU ( $V_{BU}$ ) approaches for both a non-stationary ARIMA(0,1,1) and a stationary ARIMA(1,0,1) processes at the aggregate level are calculated.

##### 4.1.2.1 Integrated moving average process order one, ARIMA(0,1,1)

The analysis is begun by deriving the  $V_{BU}$ , which is defined as follows:

$$V_{BU} = \text{Var}\left(D_t - \sum_{i=1}^N f_{i,t}\right) = \text{Var}\left(\sum_{i=1}^N d_{i,t} - \sum_{i=1}^N f_{i,t}\right) = \text{Var}\left(\sum_{i=1}^N (d_{i,t} - f_{i,t})\right) \quad (4-7)$$

By substituting (4-3) and (4-6) in (4-7) the following is given:

$$V_{BU} = \text{Var}\left(\sum_{i=1}^N \varepsilon_{i,t}\right) \quad (4-8)$$

Since  $\text{Var}(\varepsilon_{i,t}) = \sigma_i^2$  and  $\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t}) = \rho_{i,j} \sigma_i \sigma_j$ , the variance of the BU approach is:

$$V_{BU} = \sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{i,j} \sigma_i \sigma_j \quad (4-9)$$

Now the variance of the forecast error for the TD approach is derived. As discussed above, it is shown that when the subaggregate items follow an ARIMA (0,1,1) process, the

aggregate family demand also follows an ARIMA (0,1,1) process (Lütkepohl, 1984). The family aggregate process is defined as follows:

$$D_t = t\mu' + \varepsilon'_t - (1 - \beta)\varepsilon'_{t-1} + D_{t-1} \quad (4-10)$$

where  $\theta' = 1 - \beta$ .

Considering  $\theta_1 = \theta_2 = \theta_3 = \dots = \theta_N = \theta$  results in the same theta also in the aggregate demand so,  $\theta' = \theta$ . Now by considering  $\theta' = 1 - \alpha_{TD}$  and  $\theta = \theta'$ , it is obvious that the optimal smoothing constant for the aggregate demand is  $\alpha_{TD} = 1 - \theta$ , which is equal to the optimal smoothing constant for the subaggregate process.

The aggregate demand and its forecast can be expressed as a function of the error terms as following:

$$D_t = t\mu' + \varepsilon'_t + \alpha_{TD}\varepsilon'_{t-1} + \alpha_{TD}\varepsilon'_{t-2} + \dots + \alpha_{TD}\varepsilon'_1 \quad (4-11)$$

Knowing that  $\varepsilon'_t = \sum_{i=1}^N \varepsilon_{i,t}$ , the following is obtained

$$Var(\varepsilon'_t) = \sum_{i=1}^N Var(\varepsilon_{i,t}) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N Cov(\varepsilon_{i,t}, \varepsilon_{j,t}) \quad (4-12)$$

The aggregate forecast is

$$F_t = t\mu' + \alpha_{TD}\varepsilon'_{t-1} + \alpha_{TD}\varepsilon'_{t-2} + \dots + \alpha_{TD}\varepsilon'_1 \quad (4-13)$$

The variance of the TD forecast error is defined as:

$$V_{TD} = Var[D_t - F_t] \quad (4-14)$$

By substituting (4-11) and (4-13) into (4-14), the variance of TD approach is:

$$V_{TD} = Var[\varepsilon'_t] \quad (4-15)$$

By substituting (4-12) into (4-15) we have:

$$V_{TD} = \sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{i,j} \sigma_i \sigma_j \quad (4-16)$$

#### 4.1.2.2 Autoregressive moving average process order one, ARIMA(1,0,1)

In this part, the variance of forecast error of the BU approach at the aggregate level is calculated when the subaggregate items follow a stationary ARIMA(1,0,1) process. The  $V_{BU}$  can be obtained as follows:

$$\begin{aligned} V_{BU} &= Var\left(D_t - \sum_{i=1}^N f_{i,t}\right) = Var\left(\sum_{i=1}^N d_{i,t} - \sum_{i=1}^N f_{i,t}\right) \\ &= \sum_{i=1}^N Var(d_{i,t} - f_{i,t}) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N Cov(d_{i,t} - f_{i,t}, d_{j,t} - f_{j,t}) \end{aligned} \quad (4-17)$$

Subsequently, the two parts of (4-17) should to be determined: i) the variance of forecast error for subaggregate item  $i$  which is calculated in (4-33) , ii) the covariance of the forecast error between subaggregate  $i$  and  $j$ .

The covariance of the forecast error between subaggregate  $i$  and  $j$  in period  $t$ ,  $Cov(d_{i,t} - f_{i,t}, d_{j,t} - f_{j,t})$  is as follows:

$$Cov(d_{i,t} - f_{i,t}, d_{j,t} - f_{j,t}) = Cov(d_{i,t}, d_{j,t}) - Cov(d_{i,t}, f_{j,t}) - Cov(f_{i,t}, d_{j,t}) + Cov(f_{i,t}, f_{j,t}) \quad (4-18)$$

Now by substituting (P-1), (P-2), (P-3) and (P-5) in Appendix P into (4-18), the following is obtained:

$$\begin{aligned} Cov(d_{i,t} - f_{i,t}, d_{j,t} - f_{j,t}) = & \lambda_0 - \frac{\alpha_j \lambda_1}{1 - \phi_i + \alpha_j \phi_i} - \frac{\alpha_i \lambda'_1}{1 - \phi_j + \alpha_i \phi_j} + \\ & \left( \frac{1}{(1 - (1 - \alpha_i)(1 - \alpha_j))} \right) \left( \alpha_i \alpha_j \lambda_0 + \frac{\alpha_i (1 - \alpha_j) \alpha_j \lambda_1}{1 - \phi_i + \alpha_j \phi_i} + \frac{\alpha_j (1 - \alpha_i) \alpha_i \lambda'_1}{1 - \phi_j + \alpha_i \phi_j} \right) \end{aligned} \quad (4-19)$$

Finally by substituting (4-33) and (4-19) into (4-17), the variance of the forecast error of the BU approach at aggregate level is:

$$\begin{aligned} V_{BU} = & \sum_{i=1}^N \left[ \gamma_0 + \frac{\alpha_i \gamma_0}{2 - \alpha_i} + \frac{2\alpha_i (1 - \alpha_i) \gamma_1}{(2 - \alpha_i)(1 - \phi_i + \alpha_i \phi_i)} + \frac{2\alpha_i \gamma_1}{1 - \phi_i + \alpha_i \phi_i} \right] + \\ & 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[ \lambda_0 - \frac{\alpha_j \lambda_1}{1 - \phi_i + \alpha_j \phi_i} - \frac{\alpha_i \lambda'_1}{1 - \phi_j + \alpha_i \phi_j} + \right. \\ & \left. \left( \frac{1}{(1 - (1 - \alpha_i)(1 - \alpha_j))} \right) \left( \alpha_i \alpha_j \lambda_0 + \frac{\alpha_i (1 - \alpha_j) \alpha_j \lambda_1}{1 - \phi_i + \alpha_j \phi_i} + \frac{\alpha_j (1 - \alpha_i) \alpha_i \lambda'_1}{1 - \phi_i + \alpha_j \phi_i} \right) \right] \end{aligned} \quad (4-20)$$

by substituting (3-2),  $\lambda_k$  and  $\lambda'_k$  defined in (P-1) in Appendix P into (4-20) and assuming that  $\phi_1 = \phi_2 = \dots = \phi_N = \phi$ ,  $\theta_1 = \theta_2 = \dots = \theta_N = \theta$  and  $\alpha_1 = \alpha_2 = \dots = \alpha_N = \alpha$ ,  $V_{BU}$  is simplified as :

$$V_{BU} = 2 \left( \sum_{i=1}^N [\sigma_i^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right) \left[ \frac{(1 + \theta^2 + \alpha\theta - \phi(1 + \theta)^2 + \phi^2\theta(2 - \alpha))}{(2 - \alpha)(1 - \phi^2)(1 - \phi + \alpha\phi)} \right] \quad (4-21)$$

Now, the derivation of the variance of forecast error for the TD approach at the aggregate level is preceded. All subaggregate items are aggregate to produce one-step-ahead estimates at the top level based on *SES*. The  $V_{TD}$  is defined as

$$V_{TD} = \text{Var}(D_t - F_t) = \text{Var}(D_t) + \text{Var}(F_t) - 2\text{Cov}(D_t, F_t) \quad (4-22)$$

Assuming  $\phi_1 = \phi_2 = \dots = \phi_N = \phi$ , and  $\theta_1 = \theta_2 = \dots = \theta_N = \theta$ , the aggregate family demand also follows an ARIMA(1,0,1) process with the following characteristics. the aggregate series can be defined as

$$D_t = \sum_{i=1}^N d_{i,t} = \mu(1 - \phi) + \phi D_{t-1} + \hat{\varepsilon}_t - \theta \hat{\varepsilon}_{t-1}. \quad (4-23)$$

where  $\hat{\varepsilon}_t = \sum_{i=1}^N \varepsilon_{i,t}$  and  $\text{Var}(\hat{\varepsilon}_t) = \hat{\sigma}^2 = \left( \sum_{i=1}^N [\sigma_i^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right)$

$$\hat{\gamma}_k = \text{Cov}(D_t, D_{t-k}) = \begin{cases} \frac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} \hat{\sigma}^2 & k = 0 \\ \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \hat{\sigma}^2 & |k| = 1, \\ \phi \gamma_{k-1} & |k| > 1 \end{cases} \quad (4-24)$$

The evaluation of  $V_{TD}$  is begun by defining the variance of demand in (4-24). The covariance between the aggregate demand and its forecast is:

$$\begin{aligned} \text{Cov}(D_t, F_t) &= \text{Cov}(F_t, \sum_{k=1}^{\infty} \alpha_{TD} (1 - \alpha_{TD})^{k-1} D_{t-k}) = \alpha_{TD} \text{Cov}(D_t, \sum_{k=1}^{\infty} (1 - \alpha_{TD})^{k-1} D_{t-k}) = \\ &= \alpha_{TD} (\text{Cov}(D_t, D_{t-1}) + (1 - \alpha_{TD}) \text{Cov}(D_t, D_{t-2}) + (1 - \alpha_{TD})^2 \text{Cov}(D_t, D_{t-2}) + \dots), \end{aligned} \quad (4-25)$$

Then by substituting (4-24) into (3-13) the following is given:

$$\text{Cov}(D_t, F_t) = \frac{\alpha \hat{\gamma}_1}{1 - \phi + \alpha \phi}. \quad (4-26)$$

Finally, the variance of forecasts can be calculated as:

$$\begin{aligned} \text{Var}(F_t) &= \text{Var}(\alpha_{TD}F_{t-1} + (1 - \alpha_{TD})F_{t-1}) \\ &= \alpha_{TD}^2 \text{Var}(D_{t-1}) + (1 - \alpha_{TD})^2 \text{Var}(F_{t-1}) + 2\alpha_{TD}(1 - \alpha_{TD})\text{Cov}(D_{t-1}, F_{t-1}). \end{aligned} \quad (4-27)$$

By considering the fact that the process is stationary, it is clear that  $\text{Var}(F_t) = \text{Var}(F_{t-k})$  and  $\text{Cov}(D_t, F_t) = \text{Cov}(D_{t-k}, F_{t-k})$  for all  $k$  and by substituting (4-24) and (3-14) into (3-15) and then by substituting  $\hat{\sigma}^2 = \left( \sum_{i=1}^N [\sigma_i^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right)$ , the following is obtained:

$$\text{Var}(F_t) = \frac{\alpha_{TD}(1 - 2\phi\theta + \theta^2)}{(2 - \alpha_{TD})(1 - \phi^2)} + \frac{2\alpha_{TD}(1 - \alpha_{TD})(\phi - \theta)(1 - \phi\theta)}{(2 - \alpha_{TD})(1 - \phi + \alpha_{TD}\phi)(1 - \phi^2)}, \quad (4-28)$$

Finally by considering  $\hat{\gamma}_0 = \text{Var}(D_t)$  and substituting (3-14) and (3-16) into (4-22), the variance of forecast error for TD approach is obtained:

$$V_{TD} = 2 \left( \sum_{i=1}^N [\sigma_i^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right) \left[ \frac{(1 + \theta^2 + \alpha_{TD}\theta - \phi(1 + \theta)^2 + \phi^2\theta(2 - \alpha_{TD}))}{(2 - \alpha_{TD})(1 - \phi^2)(1 - \phi + \alpha_{TD}\phi)} \right] \quad (4-29)$$

### 4.1.3 Variance of forecast error at subaggregate level

In this sub-section, the variance of the forecast error consistent with the TD ( $V_{TD}$ ) and the BU approach ( $V_{BU}$ ) for the stationary ARIMA(1,0,1) process at the subaggregate level is calculated. It should be noted that the results regarding the non-stationary ARIMA(0,1,1) compared at subaggregate level are theoretically intractable. However, in the simulation study following the theoretical analysis various assumptions are relaxed and the results for the ARIMA(0,1,1) process at both levels of comparison are presented.

#### 4.1.3.1 Autoregressive moving average process order one, ARIMA(1,0,1)

In this part, the variance of forecast error of BU approach at subaggregate level is calculated, so  $V_{BU}$  is defined as:

$$V_{BU} = \sum_{i=1}^N \text{Var}(d_{i,t} - f_{i,t}) \quad (4-30)$$

Similar to (3-14) and (3-16), the variance of forecast and the covariance between the subaggregate demand and its forecast is:

$$\text{Cov}(d_{i,t}, f_{i,t}) = \frac{\alpha_i \gamma_1}{1 - \phi_i + \alpha_i \phi_i} \quad (4-31)$$

$$\text{Var}(f_{i,t}) = \frac{\alpha_i \gamma_0}{2 - \alpha_i} + \frac{2\alpha_i(1 - \alpha_i)\gamma_1}{(2 - \alpha_i)(1 - \phi_i + \alpha_i \phi_i)} \quad (4-32)$$

Now the variance of forecast error by considering  $\gamma_{i,0} = \text{Var}(d_{i,t})$  and (4-31) and (4-32) can be obtained as follows:

$$\text{Var}(d_{i,t} - f_{i,t}) = \gamma_{i,0} + \frac{\alpha_i \gamma_{i,0}}{2 - \alpha_i} + \frac{2\alpha_i(1 - \alpha_i)\gamma_{i,1}}{(2 - \alpha_i)(1 - \phi_i + \alpha_i \phi_i)} - \frac{2\alpha_i \gamma_{i,1}}{1 - \phi_i + \alpha_i \phi_i} \quad (4-33)$$

Finally, by substituting (4-33) into (4-30) and considering this assumption that  $\theta_1 = \theta_2 = \dots = \theta_N$ ,  $\phi_1 = \phi_2 = \dots = \phi_N$  and  $\alpha_1 = \alpha_2 = \dots = \alpha_N$  the following is obtained:

$$V_{BU} = \left( \frac{2(1 - \phi)((1 + \theta^2 - 2\phi\theta) + \alpha\theta(1 + \phi))}{(2 - \alpha)(1 - \phi + \alpha\phi)(1 - \phi^2)} \right) \sum_{i=1}^N \sigma_i^2 \quad (4-34)$$

Now the variance of the forecast error of the TD approach is derived when the comparison is undertaken at subaggregate level. The variance of forecast error for the TD approach,  $V_{TD}$  is defines as follows:

$$V_{TD} = \sum_{i=1}^N \text{Var}(d_{i,t} - p_i F_t) = \sum_{i=1}^N \left( \text{Var}(d_{i,t}) + p_i^2 \text{Var}(F_t) - 2p_i \text{Cov}(d_i, F_t) \right) \quad (4-35)$$

The covariance between subaggregate items  $i$  and aggregate forecast in period  $t$  is:

$$\text{Cov}(d_{i,t}, F_t) = \text{Cov}\left(d_{i,t}, \sum_{k=1}^{\infty} \alpha_{TD} (1 - \alpha_{TD})^{k-1} D_{t-k}\right), \quad i = 1, 2, \dots, N \quad (4-36)$$

By substituting  $D_{t-k} = \sum_{i=1}^N d_{i,t-k}$  into (4-36) and assuming that  $\phi_1 = \phi_2 = \dots = \phi_N = \phi$ , and  $\theta_1 = \theta_2 = \dots = \theta_N = \theta$ , the value of  $\text{Cov}(d_{i,t}, F_t)$  is derived through recursive substitutions. Recall that  $\text{Cov}(\varepsilon_{i,t-k}, \varepsilon_{i,t-k}) = \sigma_i^2$ ,  $\text{Cov}(\varepsilon_{i,t-k}, \varepsilon_{j,t-k}) = \rho_{ij} \sigma_i \sigma_j$ ,  $\text{Cov}(\varepsilon_{i,t}, \varepsilon_{i,t-k}) = 0$ , for all  $k > 0$  and  $\text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t-k}) = 0$ , for all  $k > 0$

$$\text{Cov}(d_{i,t}, F_t) = \frac{\alpha_{TD} (\phi - \theta) (1 - \phi\theta)}{(1 - \phi^2) (1 - \phi + \alpha_{TD} \phi)} \left( \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j \right), \quad i = 1, 2, \dots, N \quad (4-37)$$

Now, by substituting (3-2), (3-16) and (4-37) into (4-35) the following is given:

$$V_{TD} = \frac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} \sum_{i=1}^N \sigma_i^2 + \frac{v}{(2 - \alpha_{TD})(1 - \phi + \alpha_{TD} \phi)(1 - \phi^2)}$$

where

$$(4-38)$$

$$v = \left[ \alpha_{TD} \left( (1+\phi)(1-\theta)^2 \right) - \alpha_{TD}^2 (\phi + \phi\theta^2 - 2\theta) \right] \left( \sum_{i=1}^N p_i^2 \right) \left( \sum_{i=1}^N [\sigma_i^2] + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right) \\ - \left[ 2\alpha_{TD} (2 - \alpha_{TD}) (\phi - \theta) (1 - \phi\theta) \right] \sum_{i=1}^N p_i \sigma_i \left( \sum_{j=1}^N \rho_{ij} \sigma_j \right)$$

## 2. Theoretical Comparison

In this section, the conditions under which each approach outperforms the other one are analytically identified. The ratio of the variance of forecast error corresponding to the TD approach ( $V_{TD}$ ) to the variance of the forecast error associated with the BU approach ( $V_{BU}$ ) is calculated. A ratio that is lower than one, implies a benefit in favour of the TD approach. Conversely, if the ratio is greater than one, then the BU approach performs better (and if the ratio is equal to one, both strategies perform the same).

### 4.2.1 Comparison at aggregate level

In this sub-section, for each process under consideration the ratio of  $V_{TD}$  to  $V_{BU}$  is derived. The comparison is undertaken at the aggregate level.

#### 4.2.1.1 Integrated moving average process order one ARIMA(0,1,1)

**Proposition1.** If all the subaggregate demand items follow an ARIMA(0,1,1) process with identical moving average parameters ( $\theta_1 = \theta_2 = \theta_3 = \dots = \theta_N$ ) and the optimal smoothing constant value is used to forecast both the subaggregate and aggregate demand, then the performance of the TD and BU approaches for forecasting aggregate demand is identical ( $V_{TD} = V_{BU}$ ).

**Proof:**

The effectiveness of the TD and the BU approaches can be compared by evaluating the ratio of the corresponding variances of forecast error (i.e. by dividing (4-9) and (4-16)):

$$\frac{V_{TD}}{V_{BU}} = \frac{\left( \sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{i,j} \sigma_i \sigma_j \right)}{\left( \sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{i,j} \sigma_i \sigma_j \right)} = 1 \quad (4-39)$$

#### 4.2.1.2 Autoregressive moving average process order one ARIMA(1,0,1)

The ratio of the  $V_{TD}$  to  $V_{BU}$  when the subaggregate demand items follow an ARIMA(1,0,1) process is obtained by dividing (4-21) to (4-29) :

$$\frac{V_{TD}}{V_{BU}} = \frac{(2-\alpha)(1-\phi+\alpha\phi)(1+\theta^2+\alpha_{TD}\theta-\phi(1+\theta)^2+\phi^2\theta(2-\alpha_{TD}))}{(2-\alpha_{TD})(1-\phi+\alpha_{TD}\phi)(1+\theta^2+\alpha\theta-\phi(1+\theta)^2+\phi^2\theta(2-\alpha))} \quad (4-40)$$

This ratio is a function of the moving average parameter ( $\theta$ ), the autoregressive parameter ( $\phi$ ), and the smoothing constants ( $\alpha$  and  $\alpha_{TD}$ ). From (4-21) and (4-29) it is obvious that the optimal values of  $\alpha$  and  $\alpha_{TD}$  are equal. Hence, both  $V_{BU}$  and  $V_{TD}$  can be minimized by having the equal value of  $\alpha$  and  $\alpha_{TD}$ .

**Proposition 2:** If the time series of the all sub-aggregate demand follows an ARIMA(1,0,1) process when  $\phi_1 = \phi_2 = \dots = \phi_N$  and  $\theta_1 = \theta_2 = \dots = \theta_N$ , both the TD and the BU strategies perform equally as long as the smoothing constants used for forecasting the subaggregate demands and the aggregate demand are set optimal.

**PROOF:** By substituting ( $\alpha = \alpha_{TD}$ ) in (4-40), it is easy to demonstrate that  $V_{TD}/V_{BU} = 1$ .

These findings are in agreement with the results reported by Widiarta et al. (2009) which theoretically shows that there is no significant difference between the TD and BU approaches on forecasting aggregate demand when all subaggregate items follow an MA(1) process with identical process parameters.

## 4.2.2 Comparison at subaggregate level

In this part, the variance of the forecast error provided by the BU and the TD approaches are compared at the subaggregate level when the subaggregate demands follow an ARIMA(1,0,1) process. As explained above the comparison at the subaggregate level for the ARIMA(0,1,1) is not traceable.

### 4.2.2.1 Autoregressive moving average process order one ARIMA(1,0,1)

The ratio of  $V_{TD}$  to  $V_{BU}$  comparing at the subaggregate level is given by dividing (4-38) into (4-34). It should be noted that it is difficult to analyse the parameters with many subaggregate items, therefore the following analysis is restricted to a family with two SKUs to obtain the meaningful insights. In addition, it is assumed that  $\sigma_1 = \sigma_2$ , therefore the following is given:

$$\frac{V_{TD}}{V_{BU}} = \frac{(1 - 2\phi\theta + \theta^2) + v_R}{\left( \frac{2(1-\phi)((1+\theta^2 - 2\phi\theta) + \alpha\theta(1+\phi))}{(2-\alpha)(1-\phi + \alpha\phi)} \right)} \quad (4-41)$$

where

$$v_R = \frac{(1 + \rho_{12}) \left[ (\alpha_{TD} ((1+\phi)(1-\theta)^2) - \alpha_{TD}^2 (\phi + \phi\theta^2 - 2\theta)) (p_1^2 + p_2^2) - (\alpha_{TD} (\phi - \theta)(1 - \phi\theta)(2 - \alpha_{TD})) \right]}{(2 - \alpha_{TD})(1 - \phi + \alpha_{TD}\phi)}$$

**THEOREM 4-1:** If the time series of all subaggregate demand follows an ARIMA(1,0,1) process when  $0.5 < \phi_1 = \phi_2 < 1$  and  $-1 < \theta_1 = \theta_2 \leq 0$ , then the BU outperforms the TD approach regardless of the cross-correlation  $\rho_{12}$ , the relative weight of each subaggregate item  $p_i$ , and the smoothing constant values.

**PROOF:** Proof in Appendix Q.

**THEOREM 4-2:** If the time series of all sub-aggregate demand follows an ARIMA(1,0,1) process when i)  $-1 < \phi_1 = \phi_2 \leq .5$  and  $-1 < \theta_1 = \theta_2 < 1$  and ii) the smoothing constants used for forecasting the subaggregate demands under the BU and TD approach are set small ( $\alpha, \alpha_{TD} \leq 0.01$ ), then the maximum difference between the BU and the TD to forecast the subaggregate forecasts is 1%,  $0.99 \leq V_{TD}/V_{BU} \leq 1.01$ .

**PROOF:** Proof in Appendix R.

**COROLLARY 4.1** when the smoothing constants are set equal to 0.05, 0.15 and 0.3 in Theorem 2 above, then the ratio of  $V_{TD}/V_{BU}$  takes the values presented in Table 4-1 .

**Table 4-1:** The ratio of  $V_{TD}/V_{BU}$  for different control parameters and  $-1 < \phi_1 = \phi_2 \leq .5$ ,  $-1 < \theta_1 = \theta_2 < 1$

$\alpha = \alpha_{TD} = 0.05$	$\alpha = \alpha_{TD} = 0.15$	$\alpha = \alpha_{TD} = 0.3$
$0.95 \leq V_{TD}/V_{BU} \leq 1.1$	$0.85 \leq V_{TD}/V_{BU} \leq 1.3$	$0.72 \leq V_{TD}/V_{BU} \leq 1.6$

**COROLLARY 4.2** If the time series of all sub-aggregate demand follows an ARIMA(1,0,1) process when i)  $0.5 < \phi_1 = \phi_2 < 1$  and  $0 < \theta_1 = \theta_2 < 1$  ii) the smoothing constants are set equal to 0.01, 0.05, 0.15 and 0.3 in Theorem 5 above, then the ratio of  $V_{TD}/V_{BU}$  takes the values presented in Table 4-2.

**Table 4-2:** The ratio of  $V_{TD}/V_{BU}$  for different control parameters and  $0.5 < \phi_1 = \phi_2 < 1$ ,  $0 < \theta_1 = \theta_2 < 1$

$\alpha = \alpha_{TD} = 0.01$	$\alpha = \alpha_{TD} = 0.05$	$\alpha = \alpha_{TD} = 0.15$	$\alpha = \alpha_{TD} = 0.3$
$0.99 \leq V_{TD}/V_{BU} \leq 1.99$	$0.95 \leq V_{TD}/V_{BU} \leq 5.85$	$0.87 \leq V_{TD}/V_{BU} \leq 14.81$	$0.77 \leq V_{TD}/V_{BU} \leq 26.4$

The results of Theorem 1 show that when  $\theta$  is negative and  $\phi$  takes high positive values then the BU approach always provides more accurate forecasts than the TD one regardless of the values of the smoothing constant, the correlation between subaggregate

items, and the proportional weights. While, for the other process parameter combinations, the superiority is a function of the control and the process parameters.

When the demand follows an ARIMA(1,0,1) process, it is discussed that for the negative values of  $\theta$  and the positive values of  $\phi$ , the autocorrelation is highly positive, therefore when the autocorrelation is highly positive the BU outperforms the TD approach. When the autocorrelation is positive, successive values of  $d_t$  are positively correlated and the process will tend to be smoother than the random series. When the aggregate forecasts are disaggregate, the performance of the TD approach is deteriorated by the disaggregation process. However, the BU is not affected by that. Therefore, in these cases the BU approach outperforms the TD one.

### 3. Simulation study

In this section, a simulation study is performed to evaluate the relative performance of the TD over the BU approach under more realistic assumptions. In particular the following scenario for both the ARIMA(1,0,1) and the ARIMA(0,1,1) processes are considered. A simulation investigation is conducted to discuss the effectiveness of the BU and the TD approaches compared at the subaggregate and the aggregate level for non-identical ( $\theta_1 \neq \theta_2 \neq \dots \neq \theta_N$ ,  $\phi_1 \neq \phi_2 \neq \dots \neq \phi_N$ ) process parameters. In both approaches, the search procedure is performed in the whole range of  $-1 \leq \theta_i \leq 1$  and  $-1 \leq \phi_i \leq 1$ .

#### 4.3.1 Simulation design

The presentation of the results and the analysis of the parameters on the ratio of  $V_{TD} / V_{BU}$  becomes complex when many SKUs in the simulation experiments are considered. Therefore, the simulation analysis is restricted to a family of two SKUs to obtain the meaningful insights. This is in concordance with most of the earlier papers using simulation approaches as they have also restricted the number of items to two (Dangerfield and Morris, 1992; Flidner, 1999; Widiarta et al., 2008, 2009). The parameter values for our simulation experiment are presented in Table 1.

**Table 4-3:** Parameters of the simulation experiment

$\mu_i$	$\sigma_i^2$	$\alpha_i, \alpha_{TD}$	$\theta_i$	$\phi_i$	$\rho_{ij}$	N° Replications	N° Time Periods
400	900	0.01: 0.99	-0.9 : +0.9	-0.9 : +0.9	-0.9: +0.9	100	1000

The subaggregate demands in each period are generated randomly subject to the parameters described in **Table 4-3**. The value of  $\sigma_i$  is set quite smaller than  $\mu_i$  to avoid the generation of negative subaggregate values. Experiments have also been conducted with other values of  $\mu_i$  and  $\sigma_i$  but they are not reported here as they lead to the same insights.

To generate the demands in each period  $t$ , the error terms  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  with a cross-correlation coefficient of  $\rho_{12}$  are first generated randomly then the equations (4-1) and (3-7) are used to generate the correlated subaggregate demands. The generated demand is initialized at the value of the mean plus an error term. The simulation experiment is designed and run in Matlab 7.10.0. For each parameter combination described in Table 4 demand series of 1000 observations is generated and 100 replications are introduced.

The generated demand is split for each series at both the subaggregate and aggregate level, into three parts. The first part (within sample) consists of 200 time periods and is used in order to initialise the estimates. The second part containing 250 periods is used to determine the optimal smoothing constant (i.e. the smoothing constant used in the estimation procedure that minimises the mean square error - MSE). The search procedure to find the smoothing constant that minimises the MSE is performed in the whole range [0,1], with a step increase equal to 0.01. A grid search to minimise the  $\alpha$  is conducted, however we don't use a continuous optimisation as this is not the main focus of our work and the sensitivity to the  $\alpha$  value is not that high. Note that for the BU approach, the smoothing constants are optimized for each item individually. Finally, in order to evaluate the performance of the two forecasting approaches, the value of the variance of the forecast error for the last 550 periods of the simulation (out-of-sample) is calculated. It should be noted that the initialization data of each series have been used to calculate the proportion  $p_i$  which is used to disaggregate the aggregate forecast.

The relative benefit of one forecasting approach over the other is measured by  $V_{TD}/V_{BU}$ . As previously discussed, a ratio lower than one implies that the TD approach outperforms the BU one whereas a ratio greater than one implies the opposite.

### 4.3.2 Simulation results

In this sub-section, the results of simulation study are presented when the comparison is undertaken at both subaggregat and aggregate level.

#### 4.3.2.1 Comparison at the Aggregate Level

First, the relative performance of the BU and TD approaches at the aggregate level is analysed when the subaggregate process parameters are not necessarily identical. For each experiment, the ratio of the variance of forecast error is calculated as  $Var(D_t - F_t)/Var(D_t - \sum_{i=1}^2 f_{i,t})$ .

The simulation results show that when the process parameters are identical there is no difference between the BU and the TD approach for both the ARIMA(1,0,1) and the ARIMA(0,1,1) processes. Whereas, when the process parameters are not identical, which is more realistic, the results are different.

**Figure 4-3** presents the relative performance of the BU and the TD approaches at the aggregate level forecasting when the subaggregate demand items follow an ARIMA(0,1,1) process with different values of the moving average parameter (*i.e.*  $\theta_1 \neq \theta_2$ ).

It is seen that as the cross-correlation coefficient changes from  $-0.9$  toward  $+0.9$  the ratio of  $V_{TD}/V_{BU}$  is being reduced. The ratio is higher than or equal to one, when the cross-correlation is negative, when it equals zero, and when it takes low positive values. However, the ratio is smaller than one only if the cross-correlation is (highly) positive.

The detailed results show that when the moving average parameters,  $\theta_1$  and  $\theta_2$ , take negative values (High positive autocorrelation), the performance of the BU and the TD approaches is always identical regardless of the values of the cross-correlation.

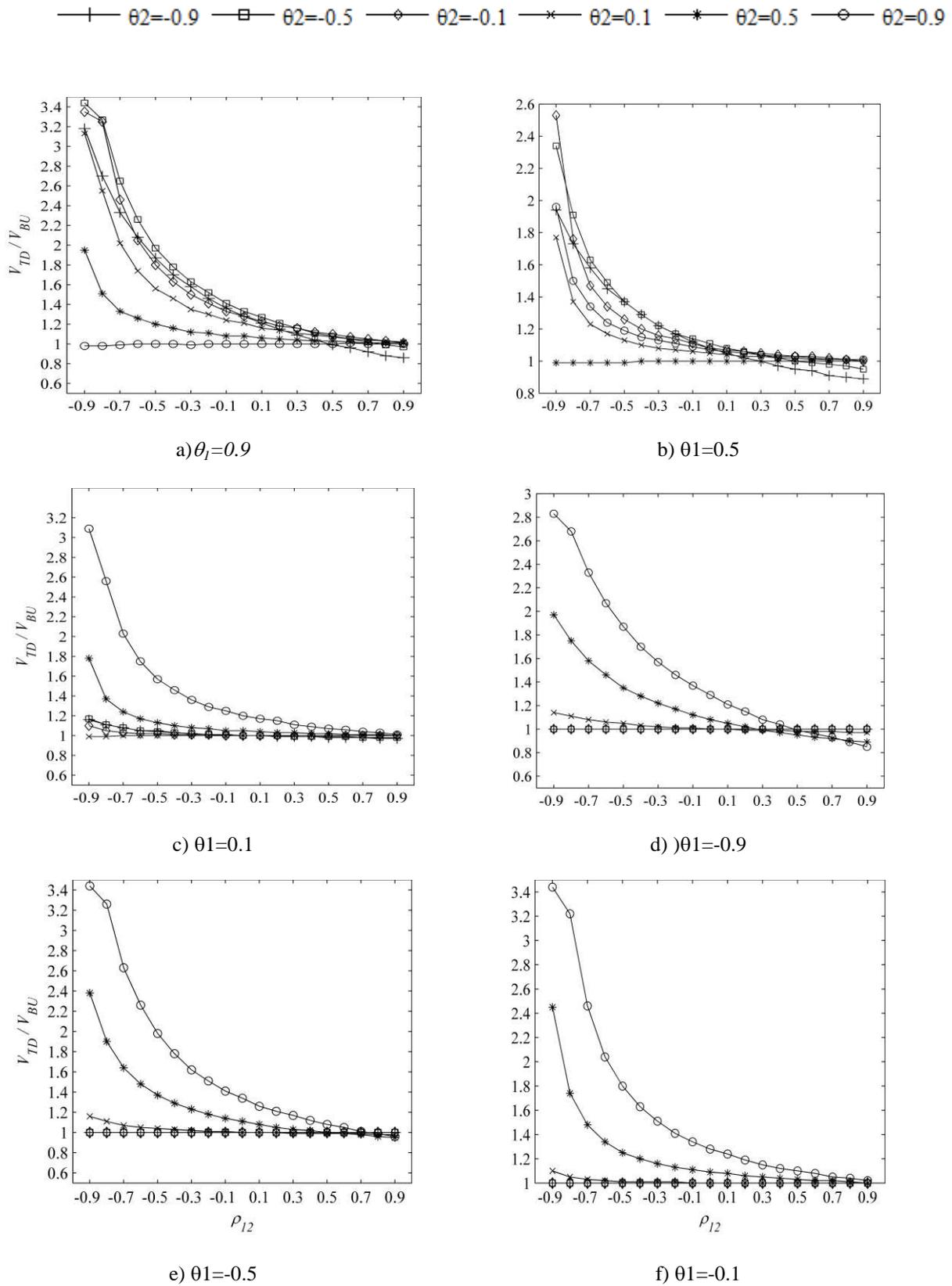
When the cross-correlation is positive the superiority of each approach depends on the value and the sign of the moving average parameters,  $\theta_1$  and  $\theta_2$ . The TD approach outperforms the BU one only when the cross-correlation is (highly) positive and the moving average parameters take high values and have opposite signs, i.e. either  $\theta_1 < 0$  and  $\theta_2 > 0$  or  $\theta_1 > 0$  and  $\theta_2 < 0$ . Note that as the cross-correlation decreases the superiority of the TD approach decreases too. For less positive cross-correlation the ratio of  $V_{TD}/V_{BU}$  becomes equal or greater than one which means that BU is preferable. In these cases TD outperforms BU with a forecast error variance reduction that can go up to 15% when the cross-correlation is very high. By decreasing the cross-correlation to 0.5, the maximum benefit of the TD approach decreases to 5% and it tends toward zero when the cross-correlation tends towards zero as well. However, under a negative cross-correlation, the BU outperforms the TD approach.

When the  $\theta_1$  and  $\theta_2$  values are positive, the ratio is almost equal to one for high positive cross-correlation and greater than one for less positive and negative cross-correlation. In the latter case the ratio of  $V_{TD}/V_{BU}$  is increased as  $\theta_1$  takes low values and  $\theta_2$  is high and vice versa.

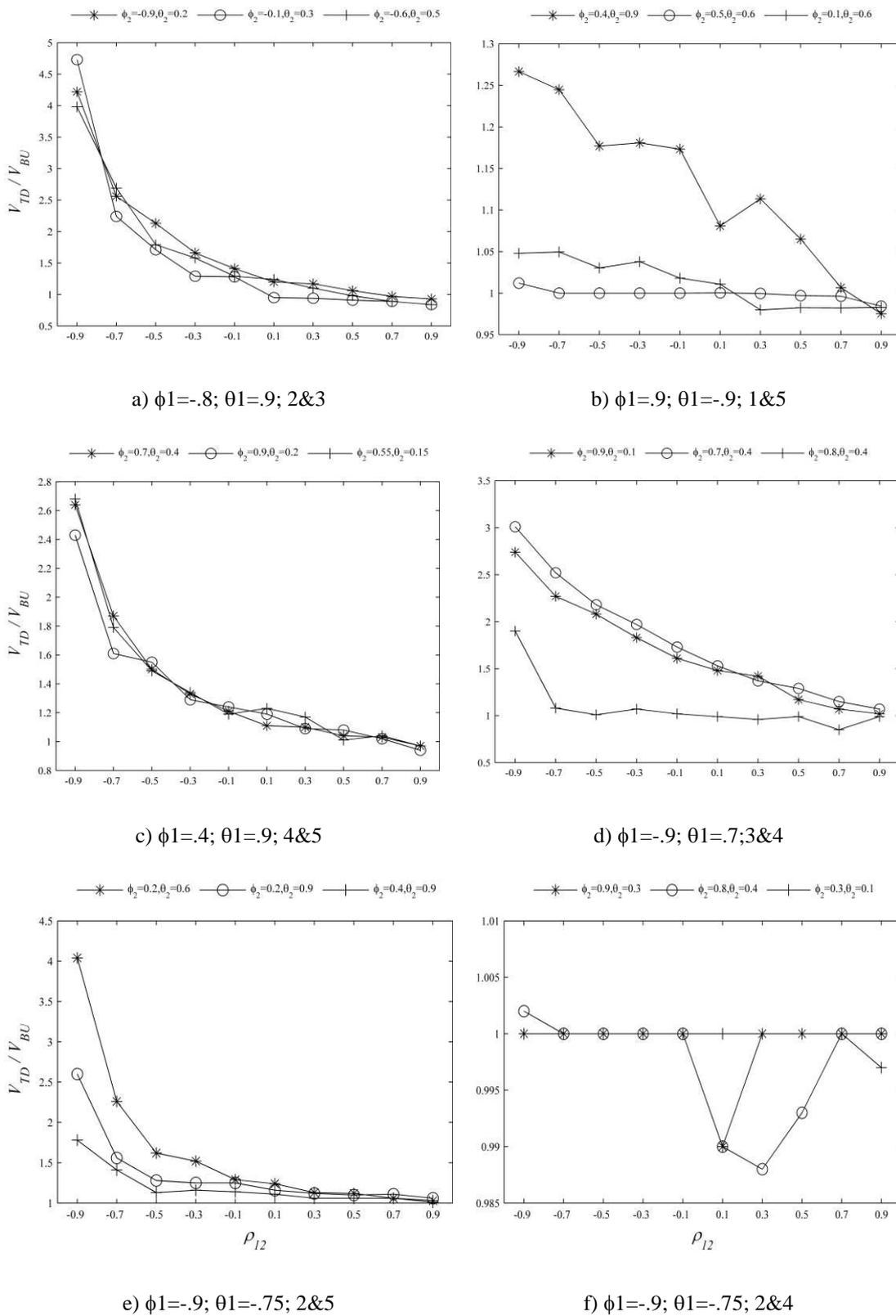
**Figure 4-4** presents the effect of the BU and the TD approaches on the demand forecasting in the aggregate level (top) when the subaggregate items follow an ARIMA(1,0,1) process with different values of the moving average and the autoregressive parameter (*i.e.*  $\theta_1 \neq \theta_2$ ,  $\phi_1 \neq \phi_2$ ).

The results show that as the cross-correlation coefficient moves from  $-0.9$  toward  $+0.9$  the ratio of  $V_{TD}/V_{BU}$  is reduced as well. The ratio is always higher than or equal to one when the cross-correlation is negative, when it equal zero, and when it takes low positive values. Thus, for these cases the BU approach provides more accurate forecasts.

The ratio may become smaller than one only if the cross-correlation is highly positive. In this case, the superiority is a function of the moving average and the autoregressive parameters. Therefore, the TD approach may outperform the BU approach when the cross-correlation is highly positive.



**Figure 4-3:** Relative performance of the TD and the BU approaches in forecasting aggregate demand under different combinations of  $\theta_1$ ,  $\theta_2$  and  $\rho_{12}$  for an ARIMA(0,1,1) demand process.



**Figure 4-4 :** Relative performance of the TD and the BU approaches in forecasting aggregate demand under different combinations of  $\theta_1$ ,  $\theta_2$  and  $\rho_{12}$  for an ARIMA(1,0,1) process.

The results show that when two subaggregate items take high positive autocorrelation, the ratio is almost equal to one regardless of the values of the cross-correlation. For example, when  $\phi_1=0.9, \theta_1=-0.8$  and  $\phi_2=-.8, \theta_2=-0.3$  ( case 1 in table 1) and  $\phi_1=0.4, \theta_1=-0.1,$  and  $\phi_2=0.9, \theta_2=0.3,$  ( case 1 and 4 in table 1). However, when two subaggregate items take opposite autocorrelation values, one with high positive and the other with negative autocorrelation, the ratio may become smaller than one and consequently the TD outperforms the BU approach. For instance when  $\phi_1=0.4, \theta_1=0.9$  and  $\phi_2=0.55, \theta_2=0.15$  ( case 4 and 5 in table 1), the forecast error variance reduction can go up to 9% when the cross-correlation is very high. This is also true when  $\phi_1=0.8, \theta_1=-0.9$  and  $\phi_2=0.1, \theta_2=0.6$ ( case 4 and 1 in table 1) for this case the variance of the forecast error reduction may go up to 3%.

In both stationary and non-stationary cases, when both subaggregate items take high positive autocorrelation, the BU and the TD approaches perform equally. One possible explanation is for a high positive autocorrelation values, the optimal value of the smoothing constant is set at the highest value in the given range which is equal to 0.99 for both TD and BU approaches. When the smoothing constant for the BU and the TD approaches is equal and the same procedure of forecasting is used, the BU and the TD approaches perform equally.

When the cross-correlation coefficient is negative, the BU approach performs better. Performance differences are further inflated when the autocorrelation values have opposite signs in which case the variance reduction achieved by the BU approach can be as high as 400% for the stationary ARIMA(1,0,1) and 500% for the non-stationary ARIMA(0,1,1) for highly negative cross-correlation. For negative cross-correlation, the pair of series moves in the opposite direction (i.e. if one increases the other decreases), therefore the subaggregate demand series have different patterns of evolution. A combination of different patterns of variation and an opposite autocorrelation values leads to a large forecast error for the TD approach and consequently large values of  $V_{TD} / V_{BU}$  for high negative cross-correlation. In these cases it is better to forecast subaggregate requirements separately and then aggregate them to get the aggregate forecast.

When the two moving average parameters take opposite signs under both processes, this means that one series has positive autocorrelation while the other has a low autocorrelation (series with random fluctuations). In addition, when the cross-correlation is

positive there is a tendency for the pair of series to move together in the same direction, so the demand series have the same pattern. When using the TD, all subaggregate are summed up series to get an aggregate one, so the fluctuations from one series may be cancelled out by others resulting in a less random series that have a lower forecast error. Therefore TD performs better than BU when the series have the same pattern associated with different autocorrelation.

In summary, when the subaggregate items follow an ARIMA(0,1,1) process and the goal is to forecast at the aggregate demand level, then the following results are achieved: *i*) the superiority of TD and BU approaches is affected by cross-correlation and autocorrelation, *ii*) if items have different patterns of fluctuation (negative cross-correlation), the ratio of  $V_{TD}/V_{BU}$  is smaller than or close to one for lower autocorrelation values, therefore the BU approach is preferred. *iii*) if the items follow the same patterns of fluctuation (high positive cross-correlation) and they have different autocorrelation patterns, one has a very high autocorrelation while the other has a lower autocorrelation values, the TD approach may outperform the BU on, *iv*) if the autocorrelation of all items is highly positive, the performance of BU and TD is always identical, and *v*) when the autocorrelation for all items is low, BU generally dominates TD, although for highly positive cross-correlation the difference is very low.

The findings are somehow in agreement with some of the earlier studies in this area by Barnea and Lakonishok (1980) and Fliedner (1999) (although we do note that our results are not directly comparable to these studies as we analyse a non-stationary case). The analysis of Barnea and Lakonishok (1980) based on empirical analysis showed that positive cross-correlation contributes to the superiority of forecasts based on aggregate data (TD), which is also the case in our study.

Fliedner (1999) used a simulation study to compare the performance of TD and BU in forecasting aggregate series where the two subaggregate items follow an MA(1) process. He found that TD dominated BU regardless of the values of the cross-correlation coefficient. They have not reported the values of  $\theta_1$  and  $\theta_2$  used in their study, so our interpretation is that this work considered only the opposing signs for  $\theta_1$  and  $\theta_2$ . Should this be the case then these findings are in agreement with ours.

### 4.3.2.2 Comparison at Subaggregate Level

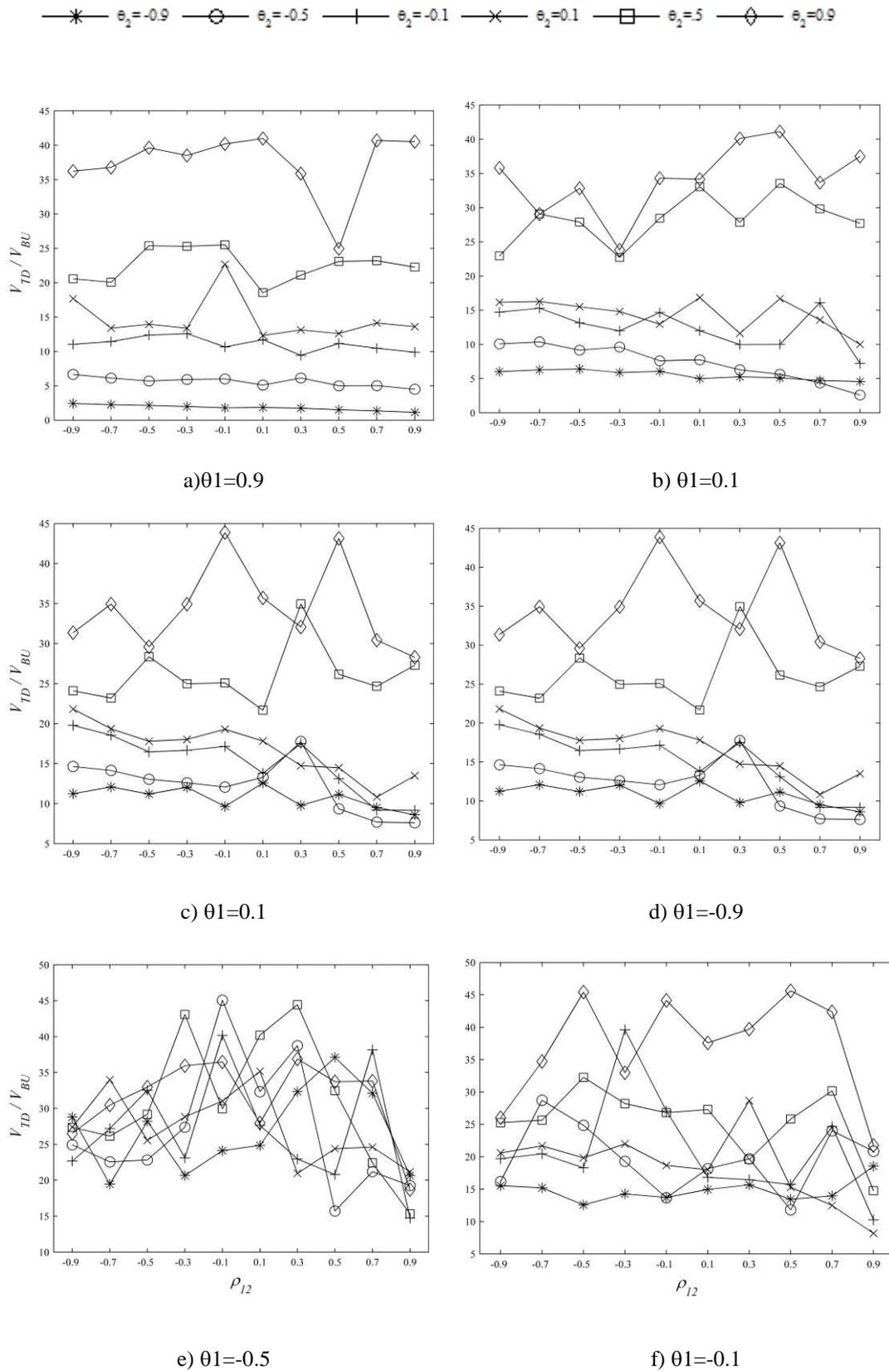
In this sub-section the relative performance of the TD and the BU approaches in forecasting subaggregate demand is evaluated when the moving average parameters are not necessarily identical. The simulation structure in terms of within and out-of-sample arrangements is as discussed in the previous sub-section. Under the BU approach, the 550 one step-ahead forecasts are generated for each item individually using the optimal smoothing constant. Under the TD approach, the sum of all subaggregate demand is calculated to obtain the aggregate series, then the aggregate forecast is provided and finally it is multiplied by the proportional contributory weight of each subaggregate item to obtain the subaggregate forecast. For each experiment, the ratio of the variance of forecast error is calculated as:

$$\sum_{i=1}^2 \text{Var}(d_{i,t} - p_i * F_t) / \sum_{i=1}^2 \text{Var}(d_{i,t} - f_{i,t}) .$$

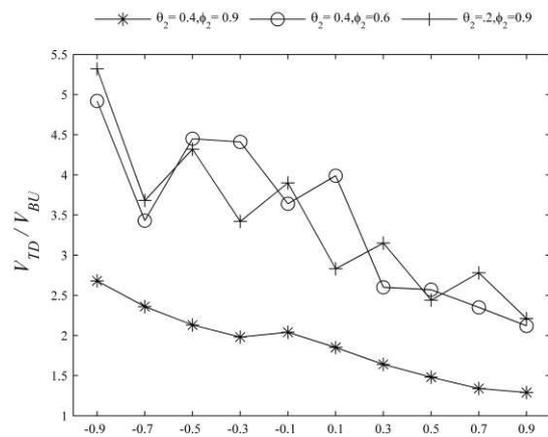
**Figure 4-5** shows the ratio of the variance of forecast error of the TD over the BU approach at the subaggregate level for different values of  $\theta_1$ ,  $\theta_2$ ,  $\rho_{12}$  when the subaggregate items follow an ARIMA(0,1,1) process with non-identical moving average parameters ( $\theta_1 \neq \theta_2$ ). The results show that when the subaggregate items follow an ARIMA(0,1,1) process, the BU approach always outperforms the TD in forecasting the subaggregate items regardless of the  $\rho_{12}$  and the process parameters.

In **Figure 4-5** it is shown that by moving from a cross-correlation of -0.9 toward +0.9 the ratio of  $V_{TD}/V_{BU}$  always remains greater than 1 regardless of the cross-correlation coefficient and the moving average parameters. When the cross-correlation and the moving average parameters,  $\theta_1$ ,  $\theta_2$ , are highly positive, *i.e.*  $\theta_1 \cong 0.99$ ,  $\theta_2 \cong 0.99$  and  $\rho_{12} \cong 0.99$ , the ratio of  $V_{TD}/V_{BU}$  becomes close to one.

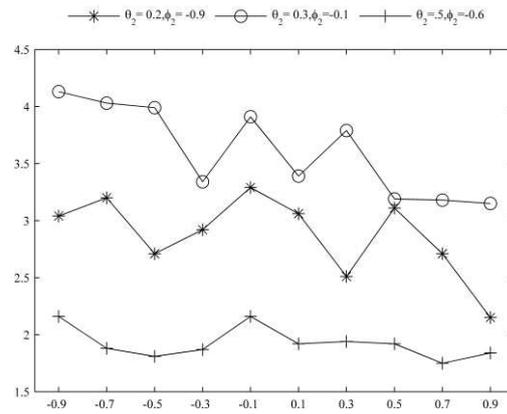
**Figure 4-5a** shows also that the BU approach outperforms the TD one by a maximum of about 80% for highly negative cross-correlation. Additionally, the rate of superiority of BU becomes very high when  $\theta_1$  and  $\theta_2$  are not highly positive (see **Figure 4-5b**, c, d).



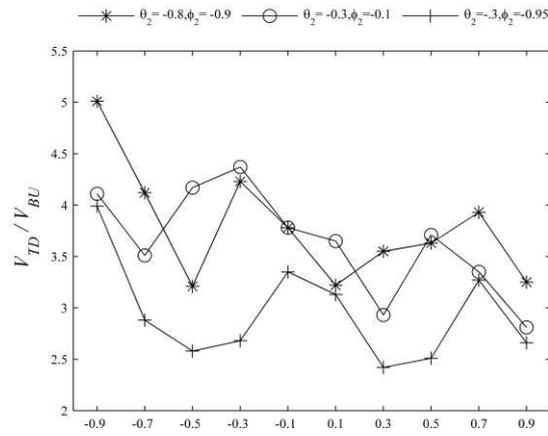
**Figure 4-5:** Relative performance of TD and BU approaches in forecasting subaggregate items under different values of  $\theta_1$ ,  $\theta_2$ ,  $\rho_{12}$  for an ARIMA(0,1,1) process.



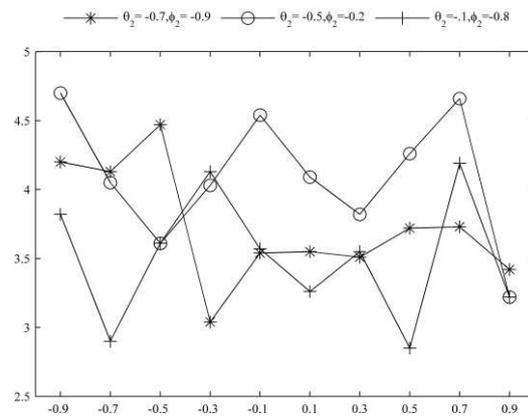
a)  $\phi_1=0.9; \theta_1=-0.7; 1&4$



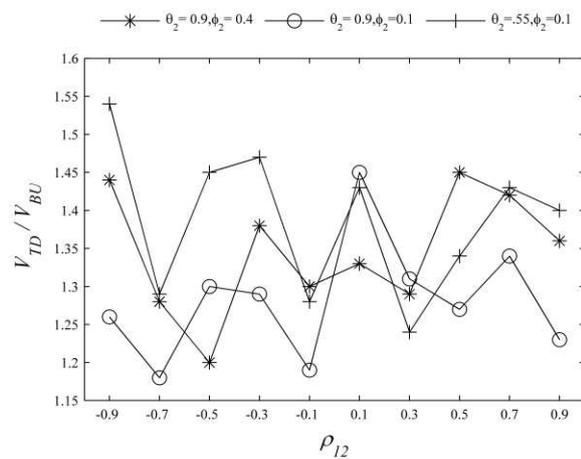
b)  $\phi_1=0.9; \theta_1=-0.9; 1&3$



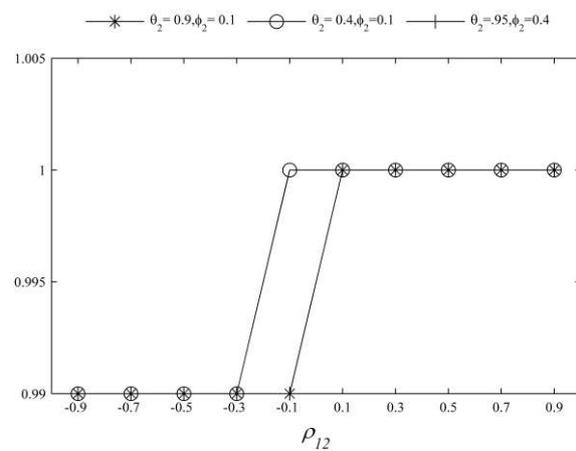
c)  $\phi_1=0.9; \theta_1=-0.85; 1&2$



d)  $\phi_1=0.9; \theta_1=0.3; 2&4$



e)  $\phi_1=0.9; \theta_1=0.4; 4&5$



f)  $\phi_1=-0.95; \theta_1=0.85; 3&5$

**Figure 4-6 :** Relative performance of TD and BU approaches in forecasting subaggregate items under different values of  $\theta_1, \theta_2, \rho_{12}$  for an ARIMA(1,0,1) process.

In **Figure 4-6** the ratio of the variance of forecast error of the TD over the BU approach is presented at the subaggregate level when the subaggregate items follow an ARIMA(1,0,1) process with non-identical moving average and autoregressive parameters ( $\theta_1 \neq \theta_2$  and  $\phi_1 \neq \phi_2$ ). The results show that the ratio of  $V_{TD}/V_{BU}$  is greater than or very close to one regardless of the values of the cross-correlation. When at least one of the subaggregate items takes high positive autocorrelation (case 1 and 4 in Table 1) the ratio is greater than one and consequently the BU approach outperforms the TD one. Additionally, by moving from high negative to high positive cross-correlation, the ratio is generally reduced. However, when none of the subaggregate items in the family take high positive autocorrelation, the difference between the BU and the TD approaches is insignificant.

The superiority of the BU at the subaggregate level can be attributed to the potentially high positive autocorrelation between demand periods. This makes it much more difficult to apportion the resulting aggregate forecast,  $Ft$ , to each item in the family based on the historical demand proportion,  $p_i$ . As a result, the performance of the TD approach is affected adversely. The performance of the BU approach, however, is not affected as it forecasts the demand for each item individually.

By comparing the results presented in **Figure 4-5** and **Figure 4-6**, it is seen that the ratio of  $V_{TD}/V_{BU}$  for the non-stationary process is much bigger than those of the stationary process. The difference of the ratio under the non-stationary ARIMA(0,1,1) and the stationary ARIMA(1,0,1) process can be attributed to the nature of these processes. When the subaggregate items follow an ARIMA(0,1,1) process, the autocorrelation is always highly positive and it spans all lags(not only lag one) except for very high positive values of the moving average parameters, however for an ARIMA(1,0,1) process the value of autocorrelation is lower and not always positive.

The findings are in accordance with those previously reported in the academic literature. Widiarta et al.(2007) argued that when the subaggregate time series follows an AR(1) process and the value of the autocorrelation is high, there is a sharp worsening in the relative performance of the TD approach. Gordon et al.(1997) and Dangerfield and Morris (1992) used the empirical data from the M-competition database and indicated that the BU dominated the TD approach when forecasting the subaggregate time series. Weatherford et

al.(2001) showed that a purely subaggregate forecast (BU) strongly outperformed even the best aggregate forecast (TD) at the subaggregate level.

These results generally confirm the findings although it must be noted that (as we mentioned in the previous sub-section) there is not a direct comparison between these studies and ours due to the consideration of a non-stationary ARIMA(0,1,1) time series process. Contrasting our results with those reported by Widiarta et al.(2007, 2009) on stationary MA(1) and AR(1) processes, it is revealed that the rates of superiority of the BU approach when the process is non-stationary is much higher than the stationary case. When the demand follows a stationary AR(1) process, the maximum ratio of  $V_{TD}/V_{BU}$  is around 6 and is obtained with series with high positive autocorrelation, while this ratio for the IMA(1,1) process is higher than 50.

## **4. Empirical analysis**

In this section, the empirical validity of the results are assess. First, the details of the empirical data available for the purposes of our investigation along with the experimental structure employed in our work are provided. Then, the results of empirical in investigation is presented.

### **4.4.1 Empirical dataset and experiment details**

The demand dataset available for the purposes of this research consists of 103 weekly sales observations (i.e. it spans a period of two years) for 1,798 SKUs from a European grocery store. The Forecast package in R is used to identify the underlying ARIMA demand process for each series and estimate the relevant parameters. It is found that more than 23% of the series (424 series) may be represented by the ARIMA(0,1,1) and more than 5% of the series (91 series) represented by ARIMA(1,0,1). It should be noted that for more than 80% of SKUs (73 SKU) the autocorrelation is relatively high positive. As such, the data does not cover the entire theoretically feasible range of the parameters. the characteristics of the SKUs relevant to this study are summarized by indicating the estimated parameters for the ARIMA(0,1,1) and ARIMA(1,0,1) process in Table 4-4 and 4-5, respectively.

**Table 4-4:** The empirical data set for ARIMA(0,1,1)

Group	$\theta$ intervals	Average of $\theta$	No. of SKUs
1	[0.1,0.3[	0.2097	4
2	[0.3,0.4[	0.3652	8
3	[0.4,0.5[	0.4656	17
4	[0.5,0.6[	0.5591	32
5	[0.6,0.7[	0.6561	67
6	[0.7,0.8[	0.7503	108
7	[0.8,0.9[	0.8467	141
8	[0.9,1]	0.9534	47
Total number of SKUs:			424

**Table 4-5:** The empirical data set for ARIMA(1,0,1)

$\theta$ intervals	$\phi$ intervals	Average of $\theta$	Average of $\phi$	Average lag1 Autocorrelation	No. of SKUs
[0.1,0.5[	[0.6,1[	0.356	0.771	0.5211	23
[0.5,0.9[	[0.6,1[	0.605	0.838	0.3260	39
[-0.2,-0.5[	[0.1,0.5[	-0.328	0.347	0.5631	29
Total number of SKUs:					91

To facilitate a clear presentation, the estimated parameters are grouped in intervals and the corresponding number of SKUs is given for each such interval. The average  $\theta$  value per interval is also presented for the process. This categorisation allows us to compare the empirical results with the theoretical findings. It should be remarked that the  $\theta$  parameter values are all positive, except for two SKUs, and most of them take highly positive values. As such, the data do not cover the entire theoretically feasible range of the parameters. The data series is divided into three parts. The first part (within sample) consists of 20 time periods and is used in order to initialize the *SES* estimates. The second part consists of 27 time periods which are used to determine the optimal smoothing constant (optimisation part); the values of the smoothing constant are varied from zero to one with a step increase of 0.01. The remaining 56 time periods are used to evaluate the performance of each approach (out-of-sample). In TD approach the aggregate forecast is disaggregate by using the proportion of each item in the family, which is calculated based on the historical demand in the initial part.

#### 4.4.2 Empirical results

The empirical results presented in **Table 4-6** are shown for the same  $\theta$  intervals. It can be seen that when the smoothing constant values are optimised for both the BU and the TD approaches, the variance ratio is greater than one regardless of whether the comparisons are undertaken at the aggregate or subaggregate level. This means that the BU approach provides more accurate both aggregate and subaggregate forecasts than the TD when demand follows an ARIMA(0,1,1) process and SES is the forecasting method. However, when the smoothing constants used for the BU and the TD approaches are equal, the ratio of  $V_{TD}/V_{BU}$  equals one in the case of aggregate demand forecasting.

**Table 4-6:** The empirical ratio of  $V_{TD}/V_{BU}$  for an ARIMA(0,1,1) process

Group	$\theta$ intervals	Comparison Level	
		Aggregate	Subaggregate
1	[0.1,0.3[	1.0032	2.0173
2	[0.3,0.4[	1.0536	1.9984
3	[0.4,0.5[	1.0097	1.8994
4	[0.5,0.6[	1.0006	1.5554
5	[0.6,0.7[	1.0356	1.3719
6	[0.7,0.8[	1.0070	1.1311
7	[0.8,0.9[	1.0158	1.0284
8	[0.9,1]	1.0403	1.0660
<b>Average</b>		<b>1.0232</b>	<b>1.5085</b>

As discussed above the moving average parameter  $\theta$ , for most SKUs considered in this research, is highly positive. More than 85% of the SKUs have a moving average parameter greater than 0.6 (see Table 4-4). In addition, the subaggregate cross-correlation coefficients between SKUs vary between -0.5 and +1; however most of them are positive.

The average of variance of forecast error reduction may be as high as 2% when the comparison is performed at the aggregate level, while 50% variance error reduction may be achieved for the comparison at the subaggregate level. By referring to the detailed results of the simulation study we see that for this range of moving average parameter values,  $0 < \theta < 1$ , the BU approach performs better than the TD at both comparison levels.

In Table 4-6 it is seen that when comparisons are undertaken at the aggregate level the ratio is close to one for all  $\theta$ ; this is confirmed by the simulation results where the moving

average parameters are positive and the cross-correlation is not highly negative (please refer to sub-section 4.3.2.1).

**Table 4-7:** The empirical ratio of  $V_{TD}/V_{BU}$  by considering the aggregation between different groups (intervals of  $\theta$  values)

Comparison Level	Group	1,	4	5	6
		2,3			
Aggregate	8	1. 2133	1. 1563	1. 0728	1. 0329
Subaggregate	8	1. 4389	1. 3201	1. 2653	1. 1175

With regards to the subaggregate level comparisons, the results show that the ratio is greater than one and is increasing by moving from higher values of  $\theta$  toward lower values. In addition for highly positive values of  $\theta$  and highly positive cross-correlation the ratio becomes close to one. In **Table 4-6** the results are presented assuming that SKUs fall within a particular interval of  $\theta$  values. In Table 4, the aggregation of items across different possible (ranges of) values is considered and the impact of the parameters on the superiority of each approach is evaluated.

To do so a category containing groups 1, 2 and 3 that includes 29 SKUs is created. This is regarded as a category with the lowest values of  $\theta$ . By moving from this category to groups 4, 5 and 6 the value of  $\theta$  increases. These groups with group 8 that represent the highest value of  $\theta$  are aggregate. The ratio of  $V_{TD}/V_{BU}$  is presented in **Table 4-7**. The results indicate that when the moving average parameters are different (Group 1,2,3 with 8) then the ratio is high, additionally as the  $\theta$  values increase (tending towards the values covered by group 8) the ratio decreases. This implies that when the groups of SKUs with low and high  $\theta$  values are aggregate, then there is a greater benefit of using the BU approach in terms of accuracy. This is exactly what is observed in the simulation results for 2 SKUs (one associated with a small and one with a high  $\theta$  value. These empirical results generally confirm

the findings of the theoretical and the simulation study when the subaggregate items follow an ARIMA(0,1,1) process.

**Table 4-8:** The empirical results for ARIMA(1,0,1)

$\theta$ intervals	$\phi$ intervals	Average of $\theta$	Average of $\phi$	Comparison Level	
				Subaggregate level	Aggregate level
[0.1,0.5[	[0.6,1[	0.356	0.771	1.23	1.031
[0.5,0.9[	[0.6,1[	0.605	0.838	1.10	1.008
[-0.2,-0.5[	[0.1,0.5[	-0.328	0.347	1.17	1.02
Average:				1.13	1.01

The empirical results for the ARIMA(1,0,1) process are presented in **Table 4-8**. It is shown that when the smoothing constant values are optimised for both the BU and the TD approaches, the variance ratio is greater than one regardless of whether the comparison is undertaken at the aggregate or subaggregate level. In addition, in the aggregate demand forecasting, the ratio of  $V_{TD}/V_{BU}$  is close to one. As it is explained above, for the moving average and autoregressive parameters values presented in Table 3-5, the autocorrelation is positive. For positive autocorrelation the difference between BU and TD approaches compared at subaggregate level is insignificant. These results generally confirm the analytical and the simulation results presented in Sub-sections 4.2.14.2.2 and Section 3 for the ARIMA(1,0,1) process.

## 5. Conclusion

In this chapter, the effectiveness of the bottom-up and top-down approaches is analytically evaluated to forecast the aggregate and the subaggregate demand when the subaggregate series follow either a first order integrated moving average ARIMA(0,1,1) or an autoregressive moving average process order one, ARIMA(1,0,1). Forecasting is assumed to be relying upon a Single Exponential Smoothing (SES) procedure and the analytical results were complemented by a simulation experiment at both the aggregate and subaggregate level as well as experimentation with an empirical dataset from a European superstore. Some empirical pieces of work discussed in section 2 confirm such a statement and provide support

for the frequency with which ARIMA(1,0,1) and/or ARIMA(0,1,1) processes are encountered in real world applications. In addition, SES is a most commonly employed forecasting procedure in industry and its application implies a non-stationary behaviour (SES is optimal for an ARIMA(0,1,1) process). In summary, the problem setting considered is a very realistic one. Analytical, simulation and empirical developments are based on the consideration of the variance of forecast error for TD and BU approaches and comparisons are undertaken at both subaggregate and aggregate level. The conditions under which one approach outperforms the other are identified.

It is found that when the subaggregate items follow an ARIMA(0,1,1) process, then BU outperforms TD to provide the subaggregate forecasts. However, to forecast the aggregate demand, the superiority of BU and TD approaches depends on the autocorrelation and cross-correlation values. For the less positive and negative cross-correlation values, BU performs better than or equally to TD. Additionally, when the cross-correlation takes high positive values, TD may outperform BU. TD works better if the cross-correlation is highly positive associated with combination of high autocorrelation vs. low autocorrelation subaggregate items. In addition, it is shown that for all identical moving average process parameter the performance of BU and TD is equal in forecasting aggregate demand. This is true as well when the smoothing constant used for all the subaggregate items and the aggregate level is set to be identical ( $\alpha = \alpha_{TD}$ ).

It is shown that, when all subaggregate items follow an ARIMA(1,0,1) process with identical moving average  $\theta$  and autoregressive  $\phi$  parameters, then the BU and TD approaches perform equally to forecast aggregate demand. However, when the process parameters are not identical, the results are different and depend on the autocorrelation and cross-correlation values. The simulation results show that for negative cross-correlation, BU approach provides more accurate results than TD. However, by increasing the cross-correlation values, the performance of BU decreases and those of TD increases. TD approach may provide more accurate forecasts than BU for high positive cross-correlation. TD is always preferable for high positive cross-correlation associated with high vs. low autocorrelation values.

When the comparison is undertaken at subaggregate level, if there is at least one subaggregate item in the family with high positive autocorrelation, then BU outperforms TD.

However, when there is no subaggregate item in the family with high positive autocorrelation, the difference between BU and TD is insignificant.

The results of comparison at subaggregate and aggregate level for ARIMA(1,0,1) and ARIMA(0,1,1) processes are slightly different. This could be attributed to the nature of the subaggregate process. For an ARIMA(0,1,1) process, the autocorrelation is always positive, moreover for most moving average parameters it is highly positive. However, for an ARIMA(1,0,1) process, the autocorrelation spans between -1 and +1, additionally it is highly positive only for a small range of process parameters.

## Chapter 5 Conclusions and Future Research

---

In this chapter, the main contributions and conclusions of this PhD thesis are given in a concise form. Additionally, the limitations of the work are identified and further research avenues are suggested.

This chapter is divided into four sections. First, the main contributions from this PhD thesis are presented. Second, the main conclusions resulted for each aggregation approach considered are summarized. Third, the managerial insights arising from this research are discussed. Finally, the limitations and some areas of future research are considered.

The overall goal of this research project is to analyse the impact of aggregation on demand forecasting. In other words, this research discusses whether it is appropriate to use disaggregate data to generate a forecast or whether one should aggregate data first and then provide a forecast.

In order to address the above issues and meet the objectives discussed in chapter 1, the following questions have been answered:

1. Under which conditions are the forecasts resulted from the temporally aggregate data preferred over those resulted from the disaggregate data?
2. Is there any optimal aggregation level for which the aggregation approach leads to the minimum variance of the forecast error?
3. Under which conditions does the BU outperforms the TD and vice versa?
4. What is the impact of the control and the process parameters on the superiority of each approach in both temporal and cross-sectional aggregation?

In this PhD research, all of the above questions have been answered and the contributions of this thesis are summarized in the following section.

## 1. Contributions of the Thesis

### 5.1.1 Temporal aggregation

The contributions of this PhD research concerning temporal aggregation are as follows:

- The superiority conditions of the aggregation and the non-aggregation approaches are identified. The cut-off point values are determined for given values of the aggregation level and the smoothing constant associated with the original demand series. This results in some theoretical rules showing the performance of each approach at the disaggregate and aggregate level of comparison.
- The performance of the aggregation approach is generally found to improve as the aggregation level increases. The rate of improvement though, is lower for the ARIMA(1,0,1) and the ARIMA(1,0,0) processes compared to the ARIMA(0,0,1) process. In all processes, the optimal aggregation level is the highest one in any given aggregation level range.
- The performance of the aggregation approach improves as the smoothing constant value employed at the aggregate series reduces. Our analytical results show that as the level of aggregation increases, the auto-correlation of the series reduces necessitating the employment of low smoothing constant values.
- In general, it is found that for high levels of positive autocorrelation in the original series, the aggregation approach may be outperformed by the non-aggregation one:
  - when comparing at the disaggregate level and where the autocorrelation is extremely positive, (i.e. high positive values of  $\phi$  in the ARIMA(1,0,0) process or high negative values of  $\theta$  and high positive values of  $\phi$  in the ARIMA(1,0,1) process), no level of aggregation improves the forecast accuracy. Consequently, the non-aggregation approach always provides more accurate forecasts. This is an intuitive finding since at any time period the most recent

demand information is 'precious'. In such a case the disaggregate approach works better as it fully exploits such recent information.

- However, when comparison is undertaken at the aggregate level, even for extreme positive values of the autocorrelation, the aggregation approach may outperform the non-aggregation one depending on the aggregation level. For lower values of the aggregation level, the non-aggregation approach works better. Nevertheless, by increasing the aggregation level, the aggregation approach outperforms the non-aggregation one. This is because the comparison is undertaken at the aggregate level where the cumulative  $m$  step ahead forecast is required. As the aggregation level and consequently the forecast horizon increases, the forecast accuracy resulting from the non-aggregation approach deteriorates and yields to a superiority in favour of the aggregation approach.
- For low positive or negative autocorrelation values, the aggregation approach is preferred regardless of the comparison level. When the autocorrelation is negative or less positive than the recent demand information is not that crucial, and then a more long term view on demand is preferred. This can be achieved as discussed above by selecting high aggregation levels and the low smoothing constants.
- Following from the above discussion, our analysis suggests that there are shades of aggregation (at one extreme no data aggregation) and shades of responsiveness of the forecast parameters ( $\alpha$ ,  $\beta$ ). Our findings suggest that the dominant solutions are either pure white (disaggregate data and responsive parameters) or pure black (aggregate data and stable forecasting algorithms with low  $\beta$ ). This is, up to a certain extent, an expected outcome given the hypothesized stationarity but: i) it is not obvious and to the best of our knowledge has never been shown before; ii) it sheds light to the general trade-off between stable forecast parameters (low smoothing constant values) that filter noise rather effectively but fail to react to changes in demand quickly and responsive forecast parameters (relatively higher smoothing constant values) that however are noise sensitive.

### 5.1.2 Cross sectional aggregation

The main contributions regarding cross-sectional aggregation can be summarized as follows:

- When the process parameters for all subaggregate items are identical, there is no significant difference between TD and BU approaches in forecasting the aggregate level as long as the optimal smoothing constant is used for both approaches. Moreover, the TD and BU approaches perform equally when the smoothing constants used for all the subaggregate items and the aggregate demand are set identical.
- When the subaggregate items are highly auto-correlated, the BU and TD approaches perform equally regardless of the cross-correlation values.
- TD performs better than BU in providing aggregate forecasts when the cross-correlations between subaggregate items are (highly), the autocorrelation of one item is positive whereas the other one is negative.
- BU may outperform TD when considering aggregate forecasts when the subaggregate items follow different patterns of fluctuation (negative cross-correlation). The TD appears not to be very accurate when the subaggregate items consist of different patterns.
- BU outperforms TD in providing subaggregate forecasts, when the autocorrelation of at least one item in the family is positive and the smoothing constant is set to its optimal value for both approaches, regardless of the cross-correlation, the disaggregation weights, and the values of the process parameters. The degree of superiority of the BU approach for the non-stationary case is much higher compared to the stationary one when comparing at subaggregate level.
- It is found that for the negative or the less positive autocorrelation, both BU and TD approaches perform almost equally in forecasting subaggregate demand when the optimal smoothing constants are used.
- The performance of BU is generally improved as the cross-correlation decreases, moving from positive toward negative values. Whereas, the performance of TD deteriorates as the cross-correlation decreases. For highly negative cross-correlation

values BU is always preferred. This is generally true for the comparison at the aggregate and the subaggregate levels.

- The benefits achieved by BU and TD approaches for the non-stationary demand process is higher than those associated with the stationary processes in terms of the forecast accuracy.

## **2. Managerial implications**

### **5.2.1 Temporal aggregation**

Our discussions with practitioners have revealed a misconception that temporal aggregation reduces variability, something that is clearly not the case. Although it is true that the non-overlapping temporal aggregation approach reduces the coefficient of variation leading to lower uncertainty. Practitioners have also expressed concerns with regards to the intuitively appealing loss of information associated with temporal aggregation. However, this concern is conditioned to short demand histories. Should long demand series be available the loss of information resulting from aggregation is outweighed by the benefits of uncertainty reduction.

- When applying temporal aggregation, practitioners should always opt for the highest possible aggregation level. However, it is important to note that consideration of high aggregation levels is subject to data availability. Although, this progressively becomes less of an issue in modern business settings. Clearly, aggregation may not constitute a viable option when short demand histories are available. Tremendous recent developments in terms of computing storage capacity facilitate the accumulation of very lengthy series. Although, we have come across situations/companies where only a few years' data is stored. In such cases aggregation may not be further considered. Long historical data series do not only allow for the more accurate estimation of the series' components but also permit the application of temporal aggregation approaches.

- The performance of aggregation improves as the smoothing constant value employed at the aggregate series reduces. This is an important finding from a practitioner's perspective since managers may set such values conveniently low to maximize the benefits derived from the aggregation approach. The smoothing constant value after aggregation should be generally set smaller than the smoothing constant before aggregation and specific rules and cut-off points have been offered for making such decisions.
- For high levels of positive autocorrelation in the original series, the non-aggregation approach outperforms the aggregation one in disaggregate level forecasting. This is an intuitive finding since at any time the most recent demand information is so precious in that case that the disaggregate approach works better as it fully exploits such recent information. However, on the contrary, for the low positive or negative autocorrelation when the recent demand information is not that crucial then a more long term view on demand is preferable, which can be obtained as discussed above by selecting high aggregation levels and low smoothing constants. This is also an important empirical insight since managers may know what to expect (in terms of any potential gains) based on the autocorrelation levels present in their series.
- When a long range forecast is required, the forecaster should apply the aggregation approach to provide the forecast. This is because a more long term view on demand is preferable and the aggregation approach utilizes better this information. As a general rule, the farther into the future we look, the more clouded our vision becomes and the non-aggregation approach will be less accurate than aggregation one.

### 5.2.2 Cross sectional aggregation

In practice, there are many series that are hierarchically organized and can be aggregated at several different levels based on products, geography or some other features. TD and BU forecasting approaches are extremely useful towards improving the accuracy of forecasts on different levels. For instance, in S&OP (Sales and Operations Planning) process, each department requires different levels of demand forecasts that can be achieved by applying TD and BU approaches.

When the practitioners require different hierarchical level of forecasts, choosing between BU or TD approaches depends on the autocorrelation, the cross-correlation and the comparison level.

- When the practitioners require demand forecasts at the SKU level, the autocorrelation values should be considered. If there is at least one series in the family with high positive autocorrelation then it would always be preferable to use the BU approach. In addition, the BU approach performs better when the series are associated with different patterns of fluctuation (negative cross-correlation).
- However, when the autocorrelation is less positive or negative, there is no difference between using BU and TD.

When the aggregate demand forecast is required, the values of cross-correlation and autocorrelation should be calculated.

- If the subaggregate items follow the same patterns of fluctuation (high positive cross-correlation) associated with different autocorrelation values (high vs. low), then TD would be applied.
- However, when the individual items are associated with different patterns of evolution BU is preferable. Additionally, if the autocorrelation values are negative for all subaggregate items, then the BU approach should be used.
- If the autocorrelation is positive for all subaggregate items, then both BU and TD perform equally. In addition, if one uses the same value of smoothing constants for both BU and TD, then both approaches perform equally as well.

### **3. Limitations and future research**

In this section, suggestions for future research are discussed from theoretical, simulation and empirical perspectives. Throughout this research some assumptions are considered that can be relaxed in future studies.

In Chapter 3, the effect of temporal aggregation on demand forecasting was discussed. In this research the case of non-overlapping temporal aggregation is considered when the disaggregate data follow a stationary demand processes and when the Single Exponential Smoothing forecasting method is used. Given the current under-consideration of temporal aggregation in inventory forecasting software solutions and given its value as a promising uncertainty reduction time series transformation approach that this PhD has revealed, research into any of the following areas would appear to be merited:

- Expansion of the analytical evaluation discussed in this work on higher order stationary processes and more importantly on non-stationary processes is a very important issue both from an academic and practitioner perspective.
- In this study, the Autoregressive Moving Average, ARMA type processes were assumed for the demand processes. This is a relevant assumption for fast moving items. The analytical and empirical consideration of Integer ARMA (INARMA) processes offers a great opportunity for advancements in the area of aggregation. Such processes bear a considerable relevance to intermittent demands where the benefits of aggregation may be even higher due to the reduction of zero observations.
- In this work, the effect of non-overlapping temporal aggregation on demand forecasting is analysed. Another important extension can be the consideration of the overlapping temporal aggregation.
- In this research, Single Exponential Smoothing is applied as a forecasting method; one natural extension is the consideration of other popular forecasting methods.
- This study is focused on forecasting and not inventory control. The extension of the work described in this research to cover inventory/implication metrics would allow a linkage between forecasting and stock control.
- Research on more extensive datasets (as well as analysis of empirical forecasting performance on measures other than the MSE) should allow a better understanding of the difficulties and benefits associated with aggregation.

In this research, the effectiveness of BU and TD approaches is evaluated to forecast the subaggregate and aggregate level. The case of stationary and non-stationary demand processes in conjunction with the SES forecast method is considered. Naturally, there are many other avenues for further research and the following possibilities should be very important in terms of advancing the current state of knowledge in the area of cross-sectional aggregation.

- In this research demand is assumed to be structured based on ARIMA type processes. The evaluation of the BU and TD approaches when the subaggregate items follow an Integer ARMA (INARMA) processes is an interesting subject for future work.
- The interface between (and the potential of combining) temporal and cross-sectional aggregation has received minimal attention both in academia and industry and is an issue that we will explore in the next steps of our research.
- Expansion of the work discussed in this research for other popular forecasting methods such as optimal forecast method, trend exponential smoothing and damped trend exponential smoothing models is an important issue.
- Extending the analysis in this research to consider  $n$  levels hierarchical structures would be an interesting development.

Finally, consideration of more extensive empirical datasets that cover the whole range of the process parameters should allow a better understanding of the benefits of each approach.

## References

---

- Acar, Y., Gardner, E.S., 2012. Forecasting method selection in a global supply chain. *International Journal of Forecasting* 28, 842-848.
- Aigner, D.J., Goldfeld, S.M., 1974. Estimation and Prediction from Aggregate Data when Aggregates are Measured More Accurately than Their Components. *Econometrica* 42, 113-134.
- Amemiya, T., Wu, R.Y., 1972. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* 67, 628-632.
- Anderson, O.D., 1975. On a lemma associated with Box, Jenkins and Granger. *Journal of Econometrics* 3, 151-156.
- Athanasopoulos, G., Hyndman, R.J., Song, H., Wu, D.C., 2011. The tourism forecasting competition. *International Journal of Forecasting* 27, 822-844.
- Babai, M.Z., Ali, M.M., Nikolopoulos, K., 2012. Impact of temporal aggregation on stock control performance of intermittent demand estimators: Empirical analysis. *OMEGA* 40, 713-721.
- Barnea, A., Lakonishok, J., 1980. An Analysis of The Usefulness of Disaggregated Accounting Data For Forecasts of Corporate Performance. *Decision Sciences* 11, 17-26.
- Bartezzaghi, E., Verganti, R., Zotteri, G., 1999. Measuring the impact of asymmetric demand distributions on inventories. *International Journal of Production Economics* 60-61, 395-404.
- Bonomo, C., 2003. Forecasting from the Center of the Supply Chain. *Journal of Business Forecasting*, 3-9.
- Box, G., Jenkins, G.M., Reinsel, G., 2008. *Time Series Analysis: Forecasting & Control*. John Wiley and sons.
- Bozos, K., Nikolopoulos, K., 2011. Forecasting the value effect of seasoned equity offering announcements. *European Journal of Operational Research* 214, 418-427.
- Brannas, K., Hellstrom, J., Nordstrom, J., 2002. A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting* 18, 19-30.
- Brewer, K.R.W., 1973. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics* 1, 133-154.

- Chen, A., Blue, J., 2010. Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands. *International Journal of Production Economics* 128, 586-602.
- Chen, A., Hsu, C.H., Blue, J., 2007. Demand planning approaches to aggregating and forecasting interrelated demands for safety stock and backup capacity planning. *International Journal of Production Research* 45, 2269-2294.
- Chen, F., Ryan, J.K., Simchi-Levi, D., 2000. The impact of exponential smoothing forecasts on the bullwhip effect. *Naval Research Logistics (NRL)* 47, 269-286.
- Chen, H., Boylan, J.E., 2007. Use of Individual and Group Seasonal Indices in Subaggregate Demand Forecasting. *The Journal of the Operational Research Society* 58, 1660-1671.
- Chen, H., Boylan, J.E., 2008. Empirical evidence on individual, group and shrinkage seasonal indices. *International Journal of Forecasting* 24, 525-534.
- Chien, C.-F., Chen, Y.-J., Peng, J.-T., 2008. Demand forecast of semiconductor products based on technology diffusion, *Simulation Conference, 2008. WSC 2008. Winter*, pp. 2313-2322.
- Chopra, S., Meindl, P., 2001. *Supply Chain Management*. Prentice-Hall, New Jersey.
- Croston, J.D., 1972. Forecasting and Stock Control for Intermittent Demands. *Operational Research Quarterly (1970-1977)* 23, 289-303.
- Dangerfield, B., Morris, J., 1988. An empirical evaluation of top-down and bottom-up forecasting strategies., *Proceedings of the 1988 Meeting of Western Decision Sciences Institute*, pp. 322-324.
- Dangerfield, B.J., Morris, J.S., 1992. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting* 8, 233-241.
- Dekimpe, M.G., Hanssens, D.M., 2000. Time-series models in marketing:: Past, present and future. *International Journal of Research in Marketing* 17, 183-193.
- Drost, F.C., Nijman, T.E., 1993. Temporal aggregation of GARCH processes. *Econometrica* 61, 909-927.
- Duc, T.T.H., Luong, H.T., Kim, Y.-D., 2008. A measure of bullwhip effect in supply chains with a mixed autoregressive-moving average demand process. *European Journal of Operational Research* 187, 243-256.
- Dunn, D.M., Williams, W.H., Dechaine, T.L., 1976. Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association* 71, 68-71.
- Erkip, N., Hausman, W.H., Nahmias, S., 1990. Optimal Centralized Ordering Policies in Multi-Echelon Inventory Systems with Correlated Demands. *Management Science* 36, 381-392.

Fliedner, G., 1999. An Investigation of Aggregate Variable Time Series Forecast Strategies with Specific Subaggregate Time Series Statistical Correlation. *Computers and Operations Research* 26, 1133-1149.

Gardner, E.S., Jr., 1990. Evaluating Forecast Performance in an Inventory Control System. *Management Science* 36, 490-499.

Gardner, E.S., Jr., 2006. Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting* 22, 637-666.

Gardner, E.S., Jr., McKenzie, E., 1985. Forecasting Trends in Time Series. *Management Science* 31, 1237-1237.

Gaur, V., Giloni, A., Seshadri, S., 2005. Information Sharing in a Supply Chain Under ARMA Demand. *Management Science* 51, 961-969.

Gordon, T., Dangerfield, B., Morris, J., 1997. Top-down or bottom-up: Which is the best approach to forecasting? *The Journal of Business Forecasting* 16, 13-16.

Granger, C.W.J., Morris, M.J., 1976. Time series modeling and interpretation. *Journal of the Royal Statistical Society. Series A*, 246-257.

Graves, S.C., Willems, S.P., 2000. Optimizing Strategic Safety Stock Placement in Supply Chains. *Manufacturing & Service Operations Management* 2, 68-83.

Graves, S.C., Willems, S.P., 2008. Strategic Inventory Placement in Supply Chains: Nonstationary Demand. *Manufacturing & Service Operations Management* 10, 278-287.

Gross, C.W., Sohl, J.E., 1990. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting* 9, 233-254.

Grunfeld, Y., Griliches, Z., 1960. Is aggregation necessarily bad? *Review of Economics and Statistics* 42, 1-13.

Harvey, A.C., 1993. *Time Series Models*. MIT Press.

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

Hyndman, R.J., Athanasopoulos, G., 2013. *Forecasting: principles and practice*. OTexts, Australia.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 679-688.

Kleijnen, J.P.C., Groenendaal, W., 1992. *Simulation: a statistical perspective*. John Wiley & Sons, Inc, ew York, NY, USA.

- Lapide, L., 2004. Sales and Operations Planning Part I: The Process. *The Journal of Business Forecasting*, 17-19.
- Lapide, L., 2006. Top-Down & Bottom-Up Forecasting In S&OP. *The Journal of Business Forecasting* 25, 14-16.
- Lee, H.L., P., P., Whang , S., 1997a. Bullwhip effect in a supply chain. *Sloan Management Review* 38, 93-102.
- Lee, H.L., Padmanabhan, V., Whang, S., 1997b. Information distortion in a supply chain: The bullwhip effect. *Management Science* 43, 546-558.
- Lee, H.L., So, K.C., Tang, C.S., 2000. The Value of Information Sharing in a Two-Level Supply Chain. *Management Science* 46, 626-643.
- Luiz, K.H., Pedro, A.M., Pedro, L.V.P., 1992. The Effect of Overlapping Aggregation on Time Series Models: An Application to the Unemployment Rate in Brazil. *Brazilian Review of Econometrics* 12, 223-241.
- Luna, I., Ballini, R., 2011. Top-down strategies based on adaptive fuzzy rule-based systems for daily time series forecasting. *International Journal of Forecasting* 27, 708-724.
- Lütkepohl, H., 1984. Forecasting Contemporaneously Aggregated Vector ARMA Processes. *Journal of Business & Economic Statistics* 2, 201-214.
- Mahajan, S., Desai, O., 2011. Value of Information in a Serial Supply Chain under a Nonstationary Demand Process, Working Paper Series. Indian Institute of Management (IIMB), Bangalore
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R., 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting* 1, 111-153.
- Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting* 16, 451-476.
- Martel, A., Diaby, M., Boctor, F., 1995. Multiple items procurement under stochastic nonstationary demands. *European Journal of Operational Research* 87, 74-92.
- Mohammadipour, M., Boylan, J., Syntetos, A., 2012. The Application of Product-Group Seasonal Indexes to Individual Products. *Foresight: The International Journal of Applied Forecasting*, 18-24.
- Mohammadipour, M., Boylan, J.E., 2012. Forecast horizon aggregation in integer autoregressive moving average (INARMA) models. *OMEGA* 40, 703-712.
- Nahmias, S., 1993. *Production and Operations Analysis*. Irwin Publishers, Boston, MA.
- Narasimhan, S.L., McLeavey, D.W., Billington, P., Billington, P.J., 1985. *Production Planning and Inventory Control*. Allyn and Bacon Inc., 67., Boston:.

- Nikolopoulos, K., Syntetos, A.A., Boylan, J., Petropoulos, F., Assimakopoulos, V., 2011. An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting : an empirical proposition and analysis. *Journal of the Operational Research Society* 62, 544-554.
- Orcutt, G.H., Watts, H.W., Edwards, J.B., 1968. Data aggregation and information loss. *The American Economic Review* 58.
- Quenouille, M., 1958. Discrete autoregressive schemes with varying time-intervals. *Metrika* 1, 21-27.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., Ducq, Y., 2013a. Demand forecasting by temporal aggregation. *Naval Research Logistics (NRL)* 60, 479-498.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., Ducq, Y., 2013b. Forecasting aggregate ARMA(1,1) Demands: theoretical analysis of top-down versus bottom-up, the International Conference on Industrial Engineering and Systems Management, Rabat, Morocco.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., Ducq, Y., 2013c. The Impact of Temporal Aggregation on Demand Forecasting of ARMA(1, 1) Process: Theoretical Analysis, in: Bakhtadze, N., Chernyshov, K., Dolgui, A., Lototsky, V. (Eds.), *Manufacturing Modelling, Management, and Control*. Elsevier, Saint Petersburg State University and Saint Petersburg ITMO University, St. Petersburg, Russia, pp. 1696-1701.
- Rostami-Tabar, B., Babai, M.Z., Syntetos, A., Ducq, Y., 2013d. Non-stationary Demand Forecasting by Cross-Sectional Aggregation, in: Under Review, I. (Ed.).
- Sariaslan, N., 2010. The Effect of Temporal Aggregation on Univariate Time Series Analysis, Department of Statistics. Middle East Technical University, p. 150.
- Sbrana, G., Silvestrini, A., 2013. Forecasting aggregate demand: Analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework. *International Journal of Production Economics*.
- Schluter, C., Trede, M., 2011. Estimating continuous-time income models, Working Paper 18/2011. Centre for Quantitative Economics, University of Münster, Germany.
- Schwarzkopf, A.B., Tersine, R.J., Morris, J.S., 1988. Top-down versus bottom-up forecasting strategies. *International Journal of Production Research* 26, 1833-1843.
- Shlifer, E., Wolff, R.W., 1979. Aggregation and Proration in Forecasting. *Management Science (pre-1986)* 25, 594.
- Silvestrini, A., Veredas, D., 2008. Temporal aggregation of univariate and multivariate time series models: A survey, Temi di discussione (working papers). Bank of Italy, Economic Research Department.
- Souza, L.R., Smith, J., 2004. Effects of temporal aggregation on estimates and forecasts of fractionally integrated processes: a Monte-Carlo study. *International Journal of Forecasting* 20, 487-502.

Spithourakis Georgios P., Fotios Petropoulos, Vassilios Assimakopoulos, Mohamed Zied Babai, Nikolopoulos, K., 2011. Improving the Performance of Popular Supply Chain Forecasting Techniques Supply Chain Forum: An International Journal 12, 16-25.

Spyros G. Makridakis, Steven C. Wheelwright, Hyndman, R.J., 1998. Forecasting: Methods and Applications. John Wiley & Sons.

Stram, D.O., Wei, W.W.S., 1986. Temporal aggregation in the ARIMA process. Journal of Time Series Analysis 7, 279-292.

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

Syntetos., A.A., 2001. Forecasting of Intermittent Demand. Business School, Buckinghamshire Chilterns University College, Brunel University, London, UK.

Taylor, J.W., 2003. Exponential smoothing with a damped multiplicative trend. International Journal of Forecasting 19, 715-725.

Teles, P., Wei, W.W.S., Crato, N., 1999. The use of aggregate series in testing for long memory. . Bulletin of the International Statistical Institute 3, 341– 342.

Theil, H., 1954. Linear Aggregation of Economic Relations. North-Holland Publishing Company, Amsterdam, Netherlands.

Tiao, G.C., 1972. Asymptotic behaviour of temporal aggregates of time series. Biometrika 59, 525-531.

Tunc, H., Kilic, O.A., Tarim, S.A., Eksioglu, B., 2011. The cost of using stationary inventory policies when demand is non-stationary. Omega 39, 410-415.

Viswanathan, S., Widiarta, H., Piplani, R., 2008. Forecasting aggregate time series with intermittent subaggregate components: top-down versus bottom-up forecasting. IMA Journal of Management Mathematics 19, 275-287.

Weatherford, L.R., Kimes, S.E., Scott, D.A., 2001. Forecasting for hotel revenue management: Testing aggregation against disaggregation. The Cornell Hotel and Restaurant Administration Quarterly 42, 53-64.

Wei, W.W.S., 1979. Some Consequences of Temporal Aggregation in Seasonal Time Series Models, In Zellner A. (ed.) Seasonal Analysis of Economic Time Series. US Department of Commerce, Bureau of the Census, Washington, DC.

Wei, W.W.S., 2006. Time Series Analysis: Univariate and Multivariate Methods. Addison-Wesley Publishing Company, Boston, USA.

Weiss, A.A., 1984. Systematic sampling and temporal aggregation in time series models. Journal of Econometrics 26, 271-281.

- Wemmerlov, U., Whybark, D.C., 1984. Lot-sizing under uncertainty in a rolling schedule environment. *International Journal of Production Research* 22, 467-484.
- Widiarta, H., Viswanathan, S., Piplani, R., 2007. On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics (NRL)* 54, 176-188.
- Widiarta, H., Viswanathan, S., Piplani, R., 2008. Forecasting item-level demands: an analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework. *IMA Journal of Management Mathematics* 19, 207-218.
- Widiarta, H., Viswanathan, S., Piplani, R., 2009. Forecasting aggregate demand: An analytical evaluation of top-down versus bottom-up forecasting in a production planning framework. *International Journal of Production Economics* 118, 87-94.
- Willemain, T.R., Smart, C.N., Shockor, J.H., DeSautels, P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting* 10, 529-538.
- Yehuda, G., Zvi, G., 1960. Is aggregation necessarily bad? *Review of Economics and Statistics* 42, 1-13.
- Zaffaroni, P., 2007. Contemporaneous aggregation of GARCH processes. *Journal of Time Series Analysis* 28, 521-544.
- Zhang, X., 2004. Technical Note: Evolution of ARMA Demand in Supply Chains. *Manufacturing & Service Operations Management* 6, 195-198.
- Zotteri, G., Kalchschmidt, M., 2007. A model for selecting the appropriate level of aggregation in forecasting processes. *International Journal of Production Economics* 108, 74-83.
- Zotteri, G., Kalchschmidt, M., Caniato, F., 2005. The impact of aggregation level on forecasting performance. *International Journal of Production Economics* 93-94, 479-491.

## Appendices

---

### Appendix A: The relationship of autocovariance between nonaggregate and aggregate demand

It has been shown that the autocovariance function on nonaggregate and aggregate series are related as follows:

$$\gamma'_0 = (1 + B + B^2 + \dots + B^{m-1})^2 \gamma_{mk+(m-1)} \quad (\text{A-1})$$

This form can be transformed to a matrix form as follow:

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \cdot \\ \cdot \\ \gamma'_k \end{bmatrix} = A \begin{bmatrix} \gamma_{(-(m-1))} \\ \gamma_{(-(m-1)+1)} \\ \cdot \\ \gamma_0 \\ \cdot \\ \gamma_{(mk+(m-1))} \end{bmatrix} \quad (\text{A-2})$$

where

$$A = \begin{bmatrix} C & 0_{mk} & & & & \\ 0_m & C & & & & \\ & & C & & & \\ & & & C & & \\ & & & & C & \\ 0_{mk} & & & & & C \end{bmatrix}$$

and

$0_n$ :  $1 \times n$  vector of zeros

$C$ :  $1 \times (2(m-1)+1)$  vector of  $C_i$  which is the coefficient of  $B^i$  in the polynomial  $(1 + B + \dots + B^{m-1})^e$

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \cdot \\ \cdot \\ \gamma'_k \end{bmatrix} = A \begin{bmatrix} \gamma_{(-(m-1))} \\ \gamma_{(-(m-1)+1)} \\ \gamma_{(-(m-1)+2)} \\ \cdot \\ \gamma_0 \\ \cdot \\ \gamma_{(mk+(m-1))} \end{bmatrix}$$

Because of this fact that  $\gamma_{-k} = \gamma_k$   $\gamma_{(-k)} = \gamma_{(k)}$  for all  $k$ , in

the matrix A in (A-2), the first  $m-1$  column corresponding to  $\gamma_{-(m-1)}, \gamma[-(m-1)], \dots, \gamma[-1]$  can be deleted by adding them to column corresponding to  $\gamma[m-1], \dots, \gamma[1], \dots, \gamma_1$  respectively. Therefore, the following is given:

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \cdot \\ \cdot \\ \gamma'_k \end{bmatrix} = A_m \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \cdot \\ \cdot \\ \gamma_{(mk+(m-1))} \end{bmatrix} \tag{A-3}$$

Where  $A_m$  is a modified matrix A, after deletion and adding required columns. (refer to Wei (2006)). different values of the aggregation level  $m$  is used to determine the general relationships between the autocovariance of non-aggregation and aggregation approach for

the demand process under consideration in this study including : ARIMA(1,0,1), ARIMA(0,0,1) and ARIMA(1,0,0). The calculation is begun by substituting  $m=3$ .

By substituting  $m=3$  into (A-1), the following is given:

$$\gamma'_k = (1 + B + B^2)^2 \gamma_{3k+2} = (1 + 2B + 3B^2 + 2B^3 + B^4) \gamma_{3k+2} \tag{A-4}$$

Now by considering (A-2) for the ARIMA(1,0,1) process and substituting  $m=3$  in that (A-5) is obtained:

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \end{bmatrix} = A \begin{bmatrix} \gamma_{-2} \\ \gamma_{-1} \\ \gamma_0 \\ \cdot \\ \cdot \\ \gamma_{11} \end{bmatrix} \tag{A-5}$$

Where

$$A = \begin{bmatrix} 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}$$

Then the matrix  $A_m$  can be calculated by adding and removing corresponding column:

$$A_3 = \begin{bmatrix} 3 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 \end{bmatrix}$$

Therefore, the following is given:

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \cdot \\ \cdot \\ \gamma_{11} \end{bmatrix} \tag{A-6}$$

By substituting (3-2) into (A-6) , (A-7) is obtained:

$$\begin{bmatrix} \dot{\gamma}_0 \\ \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \dot{\gamma}_3 \end{bmatrix} = A_3 \begin{bmatrix} \gamma_{(0)} \\ \gamma_{(1)} \\ \vdots \\ \cdot \\ \gamma_{(11)} \end{bmatrix} \quad \begin{aligned} \gamma'_0 &= 3\gamma_0 + 4\gamma_1 + 2\gamma_2 \\ \gamma'_1 &= \gamma_1(1 + 2\phi + 3\phi^2 + 2\phi^3 + \phi^4) \\ \gamma'_2 &= \gamma_1(\phi^3 + 2\phi^4 + 3\phi^5 + 2\phi^6 + \phi^7) \\ \gamma'_3 &= \gamma_1(\phi^6 + 2\phi^7 + 3\phi^8 + 2\phi^9 + \phi^{10}) \end{aligned} \tag{A-7}$$

$$\begin{bmatrix} \dot{\gamma}_0 \\ \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \dot{\gamma}_3 \end{bmatrix} = \begin{bmatrix} 342000000000 \\ 012321000000 \\ 000012321000 \\ 000000012321 \end{bmatrix} \times \begin{bmatrix} \gamma_{(0)} \\ \gamma_{(1)} \\ \vdots \\ \cdot \\ \gamma_{(11)} \end{bmatrix}$$

By substituting  $m=4$  into (A-1) the following is given:

$$\gamma'_k = (1 + B + B^2 + B^3)^2 \gamma_{4k+3} = (1 + 2B + 3B^2 + 2B^3 + B^4)^2 \gamma_{4k+3} \tag{A-8}$$

Now by considering (A-2) for the ARIMA(1,0,1) process and substituting  $m=4$  in that (A-9) is obtained:

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \\ \gamma'_4 \end{bmatrix} = A \begin{bmatrix} \gamma_{-3} \\ \gamma_{-2} \\ \gamma_{-1} \\ \gamma_0 \\ \cdot \\ \gamma_{19} \end{bmatrix} \tag{A-9}$$

Where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix}$$

Then calculate the matrix  $A_4$  is calculated as follwos:

$$A_4 = \begin{bmatrix} 4 & 6 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} = A \begin{bmatrix} \gamma_{(-3)} \\ \gamma_{(-2)} \\ \gamma_{(-1)} \\ \gamma_{(0)} \\ \cdot \\ \gamma_{(19)} \end{bmatrix}$$

So by substituting  $A_4$  into (A-9) , the following is obtained :

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \\ \gamma'_4 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 4 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{19} \end{bmatrix} \tag{A-10}$$

by substituting (3-2) into (A-10), the following equations are obtained:

$$\begin{aligned} \gamma'_0 &= 4\gamma_0 + 6\gamma_1 + 4\phi\gamma_1 + 2\phi^2\gamma_1 \\ \gamma'_1 &= \gamma_1(1 + 2\phi + 3\phi^2 + 4\phi^3 + 3\phi^4 + 2\phi^5 + \phi^6) \\ \gamma'_2 &= \gamma_1(\phi^4 + 2\phi^5 + 3\phi^6 + 4\phi^7 + 3\phi^8 + 2\phi^9 + \phi^{10}) \\ \gamma'_3 &= \gamma_1(\phi^{12} + 2\phi^{13} + 3\phi^{14} + 4\phi^{15} + 3\phi^{16} + 2\phi^{17} + \phi^{18}) \end{aligned} \tag{A-11}$$

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \\ \gamma'_4 \end{bmatrix} = A_4 \begin{bmatrix} \gamma_{(0)} \\ \gamma_{(1)} \\ \cdot \\ \cdot \\ \gamma_{(19)} \end{bmatrix}$$

$$\begin{bmatrix} \gamma'_0 \\ \gamma'_1 \\ \gamma'_2 \\ \gamma'_3 \\ \gamma'_4 \end{bmatrix} = \begin{bmatrix} 4642000 & 0 & 000000000000 \\ 01234321000000000000 \\ 0000012 & 3 & 432100000000 \\ 0000000 & 0 & 012343210000 \\ 0000000 & 0 & 000001234321 \end{bmatrix} \begin{bmatrix} \gamma_{(0)} \\ \gamma_{(1)} \\ \cdot \\ \cdot \\ \gamma_{(19)} \end{bmatrix}$$

By following the same procedure, the relationship between the autocovariance function of non-aggregation and aggregation process when  $m=5$  :

$$\begin{aligned} \gamma'_0 &= 5\gamma_0 + 8\gamma_1 + 6\phi\gamma_1 + 4\phi^2\gamma_1 + 2\phi^3\gamma_1 \\ \gamma'_1 &= \gamma_1(1 + 2\phi + 3\phi^2 + 4\phi^3 + 5\phi^4 + 4\phi^5 + 3\phi^6 + 2\phi^7 + \phi^8) \\ \gamma'_2 &= \gamma_1(\phi^5 + 2\phi^6 + 3\phi^7 + 4\phi^8 + 5\phi^9 + 4\phi^{10} + 3\phi^{11} + 2\phi^{12} + \phi^{13}) \\ \gamma'_3 &= \gamma_1(\phi^{10} + 2\phi^{11} + 3\phi^{12} + 4\phi^{13} + 5\phi^{14} + 4\phi^{15} + 3\phi^{16} + 2\phi^{17} + \phi^{18}) \\ \gamma'_4 &= \gamma_1(\phi^{15} + 2\phi^{16} + 3\phi^{17} + 4\phi^{18} + 5\phi^{19} + 4\phi^{20} + 3\phi^{21} + 2\phi^{22} + \phi^{23}) \\ \gamma'_5 &= \gamma_1(\phi^{20} + 2\phi^{21} + 3\phi^{22} + 4\phi^{23} + 5\phi^{24} + 4\phi^{25} + 3\phi^{26} + 2\phi^{27} + \phi^{28}) \end{aligned} \tag{A-12}$$

$$\begin{bmatrix} \dot{\gamma}_0 \\ \dot{\gamma}_1 \\ \dot{\gamma}_2 \\ \dot{\gamma}_3 \\ \dot{\gamma}_4 \\ \dot{\gamma}_5 \end{bmatrix} = \begin{bmatrix} 5\gamma_0 + 8\gamma_1 + 6\gamma_2 + 4\gamma_3 + 2\gamma_4 \\ \gamma_1 + 2\gamma_2 + 3\gamma_3 + 4\gamma_4 + 5\gamma_5 + 4\gamma_6 + 3\gamma_7 + 2\gamma_8 + \gamma_9 \\ \gamma_6 + 2\gamma_7 + 3\gamma_8 + 4\gamma_9 + 5\gamma_{10} + 4\gamma_{11} + 3\gamma_{12} + 2\gamma_{13} + \gamma_{14} \\ \gamma_{11} + 2\gamma_{12} + 3\gamma_{13} + 4\gamma_{14} + 5\gamma_{15} + 4\gamma_{16} + 3\gamma_{17} + 2\gamma_{18} + \gamma_{19} \\ \gamma_{16} + 2\gamma_{17} + 3\gamma_{18} + 4\gamma_{19} + 5\gamma_{20} + 4\gamma_{21} + 3\gamma_{22} + 2\gamma_{23} + \gamma_{24} \\ \gamma_{21} + 2\gamma_{22} + 3\gamma_{23} + 4\gamma_{24} + 5\gamma_{25} + 4\gamma_{26} + 3\gamma_{27} + 2\gamma_{28} + \gamma_{29} \end{bmatrix}$$

By continuing the

calculations, the general forms can be represented as follows:

$$\begin{aligned} \gamma'_0 &= m\gamma_0 + \gamma_1(2(m-1) + 2\phi(m-2) + 2(m-3)\phi^2 + \dots + 2\phi^{m-2}) \\ &= m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \end{aligned} \tag{A-13}$$

$$\begin{aligned} \gamma'_1 &= \gamma_1(1 + 2\phi + 3\phi^2 + \dots + m\phi^m + (m-1)\phi^{m+1} + \dots + 2\phi^{2m-2} + \phi^{2m-1}) \\ &= \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \end{aligned} \tag{A-14}$$

And finally we have for  $k > 1$  we have:

$$\gamma_k = \gamma_1(\phi^{m(k-1)} + 2\phi^{(k-1)m+1} + \dots + m\phi^{km-1} + (m-1)\phi^{km} + \dots + \phi^{(k+1)m-2}) \tag{A-15}$$

Now by considering (3-22) the following is given:

$$\frac{\gamma'_3}{\gamma'_2} = \frac{\phi'\gamma'_2}{\gamma'_2} = \phi' \tag{A-16}$$

From (A-15) the following ratio can be obtained:

$$\frac{\gamma'_3}{\gamma'_2} = \frac{\phi^{2m}\gamma_1(1 + 2\phi + \dots + m\phi^{m-1} + \dots + \phi^{2m-2})}{\phi^m\gamma_1(1 + 2\phi + \dots + m\phi^{m-1} + \dots + \phi^{2m-2})} = \phi^m \tag{A-17}$$

Now by comparing (A-16) with (A-17) the relationship between the autoregressive parameter before and after aggregation is given:

$$\phi' = \phi^m \quad (\text{A-18})$$

When the demand process follows an ARIMA(0,0,1) process, the relationship between the autocovariance function of aggregation and non-aggregation demand can be obtained as follows:

$$\begin{cases} \gamma'_0 = m\gamma_0 + 2(m-1)\gamma_1 \\ \gamma'_1 = \gamma_1 \end{cases} . \quad (\text{A-19})$$

By following the same procedure the relationship between the autocovariance function of the aggregation and the non-aggregation demand of an ARIMA(0,0,1) process can be obtained as following:

$$\gamma'_0 = \gamma_0 \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right), \quad (\text{A-20})$$

$$\gamma'_1 = \gamma_0 \left( \sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right), \quad (\text{A-21})$$

$$\phi' = \phi^m, \quad (\text{A-22})$$

and for all  $k > 1$ , we have:

$$\gamma'_k = \gamma_0 \left( \phi^{(k-1)m+1} + 2\phi^{(k-1)m+2} + \dots + m\phi^{km} + (m-1)\phi^{km+1} + \dots + \phi^{(k+1)m-1} \right). \quad (\text{A-23})$$

## Appendix B: Covariance between the disaggregate demand and aggregate forecast for ARIMA(1,0,1) process

The covariance between the disaggregate demand and the forecast of aggregate demand can be calculated as follows:

$$\begin{aligned} Cov(d_t, F_T) &= Cov(d_t, \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}) = \beta Cov(d_t, D_{T-1}) + \\ &\quad \beta(1-\beta)Cov(d_t, D_{T-2}) + \beta(1-\beta)^2 Cov(d_t, D_{T-2}) + \dots \end{aligned} \quad (B-1)$$

By substituting (3-10) into (B-1) we have:

$$\begin{aligned} Cov(d_t, F_T) &= \beta Cov(d_t, d_{t-1} + d_{t-2} + \dots + d_{t-m}) \\ &\quad + \beta(1-\beta)Cov(d_t, d_{t-m-1} + d_{t-m-2} + \dots + d_{t-2m}) \\ &\quad + \beta(1-\beta)^2 Cov(d_t, d_{t-2m-1} + d_{t-2m-2} + \dots + d_{t-3m}) + \dots \\ &= \beta(\gamma_1 + \gamma_2 + \dots + \gamma_m) + \beta(1-\beta)(\gamma_{m+1} + \gamma_{m+2} + \dots + \gamma_{2m}) \\ &\quad + \beta(1-\beta)^2(\gamma_{2m+1} + \gamma_{2m+2} + \dots + \gamma_{3m}) + \dots \end{aligned} \quad (B-2)$$

By substituting (3-2) into (B-2) and some simplifications, we have

$$\begin{aligned} Cov(d_t, F_T) &= \beta\phi\gamma_0(1 + \phi + \dots + \phi^{m-1}) \\ &\quad + \beta(1-\beta)\phi^{(m+1)}\gamma_0(1 + \phi + \dots + \phi^{m-1}) \\ &\quad + \beta(1-\beta)^2\phi^{(2m+1)}\gamma_0(1 + \phi + \dots + \phi^{m-1}) + \dots \\ &= (1 + \phi + \dots + \phi^{m-1}) \times (\beta\gamma_1 + \beta(1-\beta)\phi^m\gamma_1 + \beta(1-\beta)^2\phi^{2m}\gamma_1 + \dots) \end{aligned} \quad (B-3)$$

By doing some simple calculation we get

$$Cov(d_t, F_T) = \frac{\beta\gamma_1}{1 - \phi^m + \beta\phi^m} \times \frac{1 - \phi^m}{1 - \phi} \quad (B-4)$$

### Appendix C: Variance of the aggregate forecast for the ARIMA(1,0,1) process

The variance of the aggregate forecast can also be determined like (3-16) but with different parameters. In order to obtain the value of the variance of the forecast error, we need to calculate the covariance between the aggregate demand and its forecast, so we begin by deriving the covariance between the aggregate forecast and the demand in period  $T$ :

$$\begin{aligned} Cov(D_T, F_T) &= Cov(D_T, \sum_{k=1}^{\infty} \beta(1-\beta)^{k-1} D_{T-k}) = \beta Cov(D_T, \sum_{k=1}^{\infty} (1-\beta)^{k-1} D_{T-k}) = \\ &\beta (Cov(D_T, D_{T-1}) + (1-\beta)Cov(D_T, D_{T-2}) + (1-\beta)^2 Cov(D_T, D_{T-2}) + \dots) \end{aligned} \quad (C-1)$$

The variance of forecast after aggregation can be derives as:

$$\begin{aligned} Var(F_T) &= Var(\beta D_{T-1} + (1-\beta)F_{T-1}) = \beta^2 Var(D_{T-1}) + (1-\beta)^2 Var(F_{T-1}) \\ &\quad + 2\beta(1-\beta)Cov(D_{T-1}, F_{T-1}) \end{aligned} \quad (C-2)$$

By substituting (3-22) into (C-1) we get

$$Cov(D_T, F_T) = \frac{\beta\gamma'_1}{1-\phi' + \beta\phi'} \quad (C-3)$$

Then, By using the fact that  $Var(F_T) = Var(F_{T-k})$ ,  $Cov(D_T, F_T) = Cov(D_{T-k}, F_{T-k})$  for all  $k \geq 1$  and fact that  $Var(D_{T-k}) = \gamma'_0$  for all  $k$  (the properties of stationary process) and by substituting (C-3) into (C-2), we have

$$Var(F_T) = \frac{\beta\gamma'_0}{2-\beta} + \frac{2\beta(1-\beta)\gamma'_1}{(2-\beta)(1-\phi' + \beta\phi')} \quad (C-4)$$

### Appendix D: Covariance of the aggregate demand and non-aggregate forecast for ARIMA(1,0,1) process

The covariance between the aggregate demand and the subaggregate forecast is defined as follows:

$$\begin{aligned}
Cov(D_T, f_t) &= Cov\left(D_T, \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}\right) = Cov\left(d_t + d_{t+1} + \dots + d_{t+m-1}, \sum_{k=1}^{\infty} \alpha(1-\alpha)^{k-1} d_{t-k}\right) = \\
&Cov\left(d_t, \sum_{k=1}^{\infty} (1-\alpha)^{k-1} d_{t-k}\right) + \\
&Cov\left(d_{t+1}, \sum_{k=1}^{\infty} (1-\alpha)^{k-1} d_{t-k}\right) + \dots + Cov\left(d_{t+m-1}, \sum_{k=1}^{\infty} (1-\alpha)^{k-1} d_{t-k}\right) \\
&\alpha\left(Cov(d_t, d_{t-1}) + (1-\alpha)Cov(d_t, d_{t-2}) + (1-\alpha)^2 Cov(d_t, d_{t-2}) + \dots\right),
\end{aligned} \tag{D-1}$$

By substituting (3-2) into (D-1) we get

$$Cov(D_T, f_t) = Cov(d_t, f_t) + \alpha \begin{pmatrix} \gamma_2 + (1-\alpha)\gamma_3 + (1-\alpha)^2\gamma_4 + \dots \\ \gamma_3 + (1-\alpha)\gamma_4 + (1-\alpha)^2\gamma_5 + \dots \\ \vdots \\ \gamma_m + (1-\alpha)\gamma_{m-1} + (1-\alpha)^2\gamma_{m-2} + \dots \end{pmatrix}, \tag{D-2}$$

Now by substituting (3-2) into (D-2) we get:

$$\begin{aligned}
Cov(D_T, f_t) &= Cov(d_t, f_t) + \alpha \left( \frac{\phi\gamma_1}{(1-\phi+\alpha\phi)} + \frac{\phi^2\gamma_1}{(1-\phi+\alpha\phi)} + \dots + \frac{\phi^{m-1}\gamma_1}{(1-\phi+\alpha\phi)} \right) = \\
&Cov(d_t, f_t) + \frac{\alpha\phi\gamma_1(1-\phi^{m-1})}{(1-\phi+\alpha\phi)(1-\phi)},
\end{aligned} \tag{D-3}$$

By substituting (3-14) into (D-2) we get

$$Cov(D_T, f_t) = \frac{\alpha\gamma_1}{(1-\phi+\alpha\phi)} + \frac{\alpha\phi\gamma_1(1-\phi^{m-1})}{(1-\phi+\alpha\phi)(1-\phi)} = \frac{\alpha\gamma_1[(1-\phi)+\phi(1-\phi^{m-1})]}{(1-\phi)(1-\phi+\alpha\phi)} \quad (D-4)$$

### Appendix E: Coefficient of variation before and after aggregation for ARIMA(1,0,1)

When the non-aggregate process follows an ARIMA(1,0,1) process, we show that applying the non-overlapping temporal aggregation reduces the coefficient of variation (CV). CV is an important measure in an inventory context (Bartezzaghi et al., 1999). We show below that the CV decreases as the aggregation levels increases as well.

The coefficient of variation is defined as the ratio of the standard deviation of demand to the mean of demand, the ratio of the coefficient of variation after aggregation to that before aggregation is:

$$\frac{CV_{AA}}{CV_{BA}} = \frac{\frac{\sqrt{\gamma'_0}}{\mu'}}{\frac{\sqrt{\gamma_0}}{\mu}} \quad (E-1)$$

When the non-aggregate process follows ARIMA(1,0,1) , by substituting  $\mu' = m\mu$  and (3-23) and (3-2) into (E-1) we get

$$\frac{CV_{AA}}{CV_{BA}} = \frac{\sqrt{m(1-2\phi\theta+\theta^2)+(\phi-\theta)(1-\phi\theta)\left(\sum_{k=1}^{m-1} 2(m-k)\phi^{k-1}\right)}}{m\sqrt{(1-2\phi\theta+\theta^2)}} \quad (E-2)$$

Considering that  $-1 < \theta < 1$  and  $m \geq 2$  we can show that C-2 is smaller than 1 and by increasing  $m$ , the ratio of  $CV_{AA}/CV_{BA}$  decreases.

### Appendix F: Proof of theorem 1-3

By considering (3-49) and  $MSE_{BA}/MSE_{AA} > 1$ , the quadratic function given by (F-1) should be negative

$$\left( \begin{array}{l} \phi^m(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ - 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) + 2m(1-\phi^m)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{array} \right) \beta^2$$

$$+ \left( \begin{array}{l} (1-\phi^m)(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ + 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) \\ - 4m(1-\phi^m)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{array} \right) \beta - 2m^2(1-\phi)(1-\phi^m)\eta \tag{F-1}$$

where

$$\eta = \frac{\alpha\gamma_0}{2-\alpha} - \frac{2\alpha\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)} \tag{F-2}$$

Moreover, by investigating the sign of (F-1) we can obtain the conditions under which  $MSE_{BA}/MSE_{AA}$  is smaller, equal and greater than one. Now, we verify if the quadratic function (F-1) has real roots. To do so, we define the discriminant  $\Delta$  of (F-1) as follows

$$\Delta = \left( \begin{aligned} & \left( (1-\phi^m)(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \right) \\ & + 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) \\ & - 4m(1-\phi^m)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{aligned} \right)^2$$

$$+ 8 \left( \begin{aligned} & \phi^m(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ & - 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) \\ & + 2m(1-\phi^m)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{aligned} \right), \tag{F-3}$$

Now by using the fact that  $-1 < \theta < 1, -1 < \phi \leq 0, 0 < \alpha < 1$  and  $m \geq 2$ , the values of the  $\Delta$  can be obtained. If  $\Delta < 0$  it means (F-1) has no real roots and if  $\Delta > 0$  it means (F-1) has two real roots. It can shown that  $\Delta$  in (F-3) is always positive, therefore (F-1) has two different roots called  $\beta_1$  and  $\beta_2$ , where

$$\beta_1 = \frac{\left( \begin{aligned} & \left( (1-\phi^m)(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \right) \\ & - \left( + 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) \right) \\ & - 4m(1-\phi^m)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{aligned} \right) + \sqrt{\Delta}}{2 \left( \begin{aligned} & \phi^m(1-\phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) - 2(1-\phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) + \\ & 2m(1-\phi^m)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{aligned} \right)}, \tag{F-4}$$

$$\beta_2 = \frac{- \left[ \begin{aligned} & (1 - \phi^m)(1 - \phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ & + 2(1 - \phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) \\ & - 4m(1 - \phi^m)\gamma_1 - 2m^2\phi^m(1 - \phi)\eta + m^2(1 - \phi)(1 - \phi^m)\eta \end{aligned} \right] - \sqrt{\Delta}}{2 \left[ \begin{aligned} & \phi^m(1 - \phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ & - 2(1 - \phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) + 2m(1 - \phi^m)\gamma_1 + m^2\phi^m(1 - \phi)\eta \end{aligned} \right]}. \tag{F-5}$$

It can be shown that if  $\theta < 0$ ,  $\beta_2$  is always smaller than zero and  $\beta_1 > 0$  and if  $\theta > 0$ ,  $\beta_2$  is greater than one and  $\beta_1 > 0$ .

It is known that the sign of the (F-1) between the two roots  $\beta_1$  and  $\beta_2$  is opposite to the sign of  $A$ , where  $A$  defined in (F-6) is the sign of the coefficient of  $\beta^2$ , Otherwise it is that the same as the sign of  $A$ .

$$A = \left[ \begin{aligned} & \phi^m(1 - \phi) \left( m\gamma_0 + \gamma_1 \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \right) \\ & - 2(1 - \phi) \left( \gamma_1 \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \right) + 2m(1 - \phi^m)\gamma_1 + m^2\phi^m(1 - \phi)\eta \end{aligned} \right] \tag{F-6}$$

Now by considering  $\beta_1$ ,  $\beta_2$  and  $A$  that is positive for  $\theta < 0$  and negative for  $\theta > 0$ , the sign of (F-1) is determined. So we have

- If  $\theta < 0$ ,  $\beta_2$  is always smaller than zero. If  $0 < \beta_1$  then (F-1) is negative in the interval  $[\beta_2, \beta_1]$  and it is positive outside this interval.
- If  $\theta > 0$ ,  $\beta_2$  is greater than one and we can show that  $0 < \beta_1 < \beta_2$  thus (F-1) is positive in the interval  $[\beta_1, \beta_2]$  and it is negative outside this interval.

From the above expressions we can see that when  $\beta < \beta_1$ , (F-1) is negative, otherwise when  $\beta > \beta_1$ , it is positive and when  $\beta = \beta_1$ , (F-1) is equal to zero. Equivalently

- If  $\beta < \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is greater than one and consequently the aggregation approach outperforms the non-aggregation one.
- If  $\beta = \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is equal to one and both approaches perform equally.
- If  $\beta > \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is smaller than one and the non-aggregation approach outperforms the aggregation approach.

## Appendix G: Selection procedure for the ARIMA(1,0,1) process

Using the fact that  $-1 < \theta < 1$ ,  $-1 < \phi \leq 0$ ,  $0 < \alpha < 1$  and  $m \geq 2$  the value of the discriminant  $\Delta$  and the roots  $\beta_1$  and  $\beta_2$  can be defined by (F-3), (F-4) and (F-5) respectively.

If  $\Delta < 0$  there are no real roots for (F-1), therefore the sign of (F-1) is equivalent to the sign of  $A$  defined in (F-6). We can show that when  $\Delta < 0$ ,  $A$  is always positive, consequently (F-1) is positive which means that  $MSE_{BA}/MSE_{AA}$  is smaller than one.

If  $\Delta > 0$ , (F-1) has two different roots  $\beta_1$  and  $\beta_2$ . By investigating the sign of  $\beta_1, \beta_2$  and  $A$ , we can determine the sign of (F-1) and consequently the performance superiority of each strategy.

- If  $\beta_2 < 0$  and  $\beta_1 > 0$  then (F-1) is negative in the interval  $[\beta_2, \beta_1]$  and it is positive outside this interval.
- If  $\beta_2 > 1$ , it can be shown that  $0 < \beta_1 < \beta_2$ . (F-1) is positive in the interval  $[\beta_1, \beta_2]$  and it is negative outside this interval.
- If  $0 < \beta_2 < 1$ , it can be shown that  $\beta_1 > \beta_2$  then (F-1) is negative in the interval  $[\beta_2, \beta_1]$  and it is positive outside this interval.

By considering the above expressions, we have

- If  $\beta_2 < 0$  or  $\beta_2 > 1$ , then
  - If  $\beta < \beta_1$ ,  $MSE_{BA}/MSE_{AA} > 1$ .
  - If  $\beta > \beta_1$ ,  $MSE_{BA}/MSE_{AA} = 1$ .
  - If  $\beta = \beta_1$ ,  $MSE_{BA}/MSE_{AA} < 1$ .
- Otherwise, if  $0 < \beta_2 < 1$  then
  - If  $\beta_2 < \beta < \beta_1$ ,  $MSE_{BA}/MSE_{AA} > 1$ .
  - If  $\beta = \beta_1 = \beta_2$ ,  $MSE_{BA}/MSE_{AA} = 1$ .
  - If  $\beta < \beta_2$  and  $\beta > \beta_1$ ,  $MSE_{BA}/MSE_{AA} < 1$ .

### Appendix H: Proof of theorem 2-3, ARIMA(0,0,1)

By considering  $MSE_{BA}/MSE_{AA} > 1$  and some simplifications, the quadratic function given by (H-1) should be negative

$$(2\theta - 2m\theta)\beta^2 + (m^2\eta + m(1 + \theta^2) + 2m\theta)\beta - 2m^2\eta \quad (\text{H-1})$$

where

$$\eta = \frac{\alpha + \alpha\theta^2 + 2\alpha\theta}{2 - \alpha}. \quad (\text{H-2})$$

Moreover, by investigating the sign of (H-1) we can obtain the conditions under which  $MSE_{BA}/MSE_{AA}$  is smaller, equal and greater than one. Now, we verify if the quadratic function (H-1) has real roots. To do so, we define the discriminant  $\Delta$  of (H-1) as follows

$$\Delta = \left( (m^2\eta + m(1 + \theta^2) + 2m\theta) \right)^2 + 8(2\theta - 2m\theta)m^2\eta, \quad (\text{H-3})$$

Now we use the fact that  $-1 < \theta < 1$ ,  $0 < \alpha < 1$  and  $m \geq 2$  to obtain the values of  $\Delta$ . If  $\Delta < 0$  it means that (H-1) has no real roots and if  $\Delta > 0$  it means (H-1) has two real roots. We can show that  $\Delta$  in (H-3) is always positive, therefore (H-1) has two different roots denoted by  $\beta_1$  and  $\beta_2$ , where  $\beta_1$  is defined in (F-4) and

$$\beta_2 = \frac{\left( -(m^2\eta + m(1 + \theta^2) + 2m\theta) - \sqrt{\left( \frac{(m^2\eta + m(1 + \theta^2))}{+ 2m\theta} \right)^2 + 8(2\theta - 2m\theta)m^2\eta} \right)}{2(2\theta - 2m\theta)}. \quad (\text{H-4})$$

It can be shown that if  $\theta < 0$ ,  $\beta_2$  is always smaller than zero and  $0 < \beta_1 < 1$  or  $1 < \beta_1$  and  $\theta > 0$ ,  $\beta_2$  is greater than one and  $0 < \beta_1 < 1$  or  $1 < \beta_1$ .

It is known that the sign of the (H-1) between the two roots  $\beta_1$  and  $\beta_2$  is opposite to the sign of  $A$ , where  $A = (2\theta - 2m\theta)$  is the sign of the coefficient of  $\beta^2$ . Otherwise it is that the same as the sign of  $A$ . Now by considering  $\beta_1$ ,  $\beta_2$  and  $A$  that is positive for  $\theta < 0$  and negative for  $\theta > 0$ , we determine the sign of (H-1). So we have

- If  $\theta < 0$ ,  $\beta_2$  is always smaller than zero. If  $0 < \beta_1$  then (H-1) is negative in the interval  $[\beta_2, \beta_1]$  and it is positive outside this interval.
- If  $\theta > 0$ ,  $\beta_2$  is greater than one and we can show that  $0 < \beta_1 < \beta_2$  thus (H-1) is positive in the interval  $[\beta_1, \beta_2]$  and it is negative outside this interval.

From the above expressions we can see that when  $\beta < \beta_1$ , (H-1) is negative, otherwise when  $\beta > \beta_1$ , it is positive and when  $\beta = \beta_1$ , (H-1) is equal to zero. Equivalently

- If  $\beta < \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is greater than one and consequently the aggregation approach outperforms non-aggregation approach.
- If  $\beta = \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is equal to one and both approaches perform equally.
- If  $\beta > \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is smaller than one and the non-aggregation approach outperforms the aggregation one.

### Appendix I: Selection procedure for the ARIMA(1,0,0) process

Considering  $MSE_{BA}/MSE_{AA} > 1$  is equivalent to having the quadratic function (I-1) negative, which subsequently is equivalent to

$$\begin{aligned} & (\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)\beta^2 \\ & + \left( (1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta \right)\beta \\ & - 2m^2(1-\phi^m)(1-\phi) \end{aligned} \quad (I-1)$$

For the quadratic function given by (I-1), the value of the discriminant  $\Delta$  and the roots  $\beta_1$  and  $\beta_2$  can be defined as follows:

$$\begin{aligned} \Delta = & \left( (1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta \right)^2 \\ & + 8(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)(m^2(1-\phi^m)(1-\phi)\eta), \end{aligned} \quad (I-2)$$

$$\beta_1 = \frac{\left( - \left( (1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta \right) + \sqrt{\Delta} \right)}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}, \quad (I-3)$$

$$\beta_2 = \frac{\left( - \left( (1-\phi^m)(1-\phi)\xi_1 + 2(1-\phi)\xi_2 - 4m\phi(1-\phi^m) - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi^m)(1-\phi)\eta \right) - \sqrt{\Delta} \right)}{2(\phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta)}, \quad (I-4)$$

Where

$$\xi_1 = \left( m + \sum_{k=1}^{m-1} 2(m-k)\phi^k \right), \quad (I-5)$$

$$\xi_2 = \left( \sum_{k=1}^m k\phi^k + \sum_{k=1}^{m-1} k\phi^{2m-k} \right), \quad (I-6)$$

$$\eta = \frac{(-3\alpha\phi + \alpha + \alpha^2\phi)}{(2-\alpha)(1-\phi + \alpha\phi)}. \quad (\text{I-7})$$

We define the coefficient of  $\beta^2$  in (D-1) as follows

$$A = \phi^m(1-\phi)\xi_1 - 2(1-\phi)\xi_2 + 2m\phi(1-\phi^m) + m^2\phi^m(1-\phi)\eta. \quad (\text{I-8})$$

if  $\Delta < 0$  then the non-aggregation approach is always provides more accurate forecasts, otherwise

- If  $\beta_2 < \beta < \beta_1$  then the aggregation approach works better.
- If  $\beta = \beta_1 = \beta_2$  then both approaches are identical.
- If  $\beta > \beta_1$  and/or  $\beta < \beta_2$  then the non-aggregation approach works better.

## Appendix J: Proof of theorem 3-3 and theorem 4-3

**Case 1.** Using the fact that  $1/3 < \phi < 1$ ,  $m \geq 2$  and by considering the optimal smoothing constant,  $\alpha^* = (3\phi - 1)/2\phi$  used to calculate  $MSE_{BA}$ , we can show that the discriminant  $\Delta$  defined in (I-2) is negative, so there is no real root for (I-1). Consequently, the sign of (I-1) is the same as the sign of  $A$  defined in (I-2), we can show that the sign of  $A$  is always positive, therefore (I-1) is always positive and  $MSE_{BA}/MSE_{AA}$  is smaller than one. Hence, the non-aggregation approach always works better for the whole range of  $\beta$  and for any value of the aggregation level,  $m$ .

**Case 2.**  $-1 < \phi \leq 1/3$ . Using the fact that  $-1 < \phi \leq 1/3$ ,  $m \geq 2$  and by considering the small value of the smoothing constant before aggregation,  $\alpha^* < 0.05$ , it is straightforward to show that the discriminant  $\Delta$  defined in (I-2) is positive, so (I-1) has two different roots denoted by  $\beta_1$  and  $\beta_2$  defined in (I-3) and (I-4) respectively.

We can show that the value of  $\beta_2$  is either less than zero or greater than one. Now by considering the roots  $\beta_1$ ,  $\beta_2$  and the sign of  $A$ , where  $A$  is defined in (I-8), we can determine the sign of (I-1) and consequently show the superiority of each approach.

- If  $\beta_2 < 0$  and  $\beta_1 > 0$ , then (I-1) is negative in the interval  $[\beta_2, \beta_1]$  and it is positive outside this interval.
- If  $\beta_2 > 1$ , we can show that  $0 < \beta_1 < \beta_2$  and (I-1) is positive in the interval  $[\beta_1, \beta_2]$  and it is negative outside this interval.

Now from the above expressions we can get the following results:

- If  $\beta < \beta_1$ , then  $MSE_{BA}/MSE_{AA} > 1$ .
- If  $\beta = \beta_1$ , then  $MSE_{BA}/MSE_{AA} = 1$ .
- Otherwise,  $MSE_{BA}/MSE_{AA} < 1$ .

### Appendix K: Selection procedure for the ARIMA(1,0,1) process- Comparison at the aggregate level

By considering  $MSE_{BA}/MSE_{AA} > 1$  is equivalent to having the quadratic function (K-1) negative, which subsequently is equivalent to

$$\begin{aligned}
 & (-\phi^m \Psi) \beta^2 + (2\phi^m \Psi - \Psi(1 - \phi^m) - 2\phi^m \eta + 2\xi) \beta \\
 & + 2(1 - \phi^m)(\Psi - \eta)
 \end{aligned} \tag{K-1}$$

where

$$\Psi = \left( \begin{aligned}
 & \left( \frac{m(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} \right) + \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right) \left( \sum_{k=1}^{m-1} 2(m - k)\phi^{k-1} \right) \\
 & + \frac{m^2 \alpha}{2 - \alpha} \left( \frac{1 - 2\phi\theta + \theta^2}{1 - \phi^2} \right) + \frac{2m^2 \alpha (1 - \alpha)}{(2 - \alpha)(1 - \phi + \alpha\phi)} \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right) \\
 & - \frac{2m\alpha [(1 - \phi) + \phi(1 - \phi^{m-1})]}{(1 - \phi)(1 - \phi + \alpha\phi)} \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right)
 \end{aligned} \right)$$

$$\eta = \left( \frac{m(1 - 2\phi\theta + \theta^2)}{1 - \phi^2} + \left( \sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right) \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right) \right)$$

$$\xi = \left( \sum_{k=1}^m k\phi^{k-1} + \sum_{k=2}^m (k-1)\phi^{2m-k} \right) \left( \frac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2} \right)$$

For the quadratic function given by (K-1), the value of the discriminant  $\Delta$  and the roots  $\beta_1$  and  $\beta_2$  can be defined as follows:

$$\Delta = (2\phi^m\Psi - \Psi(1 - \phi^m) - 2\eta\phi^m + 2\xi)^2 + 8\phi^m\Psi(1 - \phi^m)(\Psi - \eta) \quad (\text{K-2})$$

$$\beta_1 = \frac{(2\phi^m\Psi - \Psi(1 - \phi^m) - 2\eta\phi^m + 2\xi) + \sqrt{\Delta}}{2\phi^m\Psi} \quad (\text{K-3})$$

$$\beta_2 = \frac{-(2\phi^m\Psi - \Psi(1 - \phi^m) - 2\eta\phi^m + 2\xi) + \sqrt{\Delta}}{-2\phi^m\Psi} \quad (\text{K-4})$$

the coefficient of  $\beta^2$  in (C-1) is defined as follows:

$$A = -\phi^m\Psi. \quad (\text{K-5})$$

If the discriminant  $\Delta < 0$ , there are no real roots for (K-1), therefore the sign of (K-1) is equivalent to the sign of  $A$ . We can show that when  $\Delta < 0$ ,  $A$  is always negative, consequently (K-1) is negative which means that  $MSE_{BA}/MSE_{AA}$  is smaller than one.

However, If  $\Delta > 0$ , (K-1) has two different roots  $\beta_1$  and  $\beta_2$ . By investigating the sign of  $\beta_1, \beta_2$  and  $A$ , we can determine the sign of (K-1) and consequently the performance superiority of each approach.

- If  $\beta_2 < 0$  and  $\beta_1 > 0$  then (K-1) is positive in the interval  $[\beta_2, \beta_1]$  and it is negative outside this interval.
- If  $\beta_2 > 1$ , we can show that  $0 < \beta_1 < \beta_2$ . (K-1) is negative in the interval  $[\beta_1, \beta_2]$  and it is positive outside this interval.
- If  $0 < \beta_2 < 1$ , we can show that  $\beta_1 > \beta_2$  then (K-1) is positive in the interval  $[\beta_2, \beta_1]$  and it is negative outside this interval.

By considering the above expressions, the superiority conditions of each approach can be obtained by following the selection procedure :

1. The procedure is begun by calculating  $\Delta$  defined in (K-2), If  $\Delta < 0$  then the non-aggregation approach is always superior, otherwise the values of  $\beta_1$  and  $\beta_2$  defined in (K-3) and (K-4) are calculated.
2. If  $\beta_2 < 0$  or  $\beta_2 > 1$ , the value of  $\beta_1$  is calculated,
  - If  $\beta < \beta_1$  then the aggregation approach works better.
  - If  $\beta = \beta_1$  then both approaches are identical.
  - If  $\beta > \beta_1$  then the non-aggregation approach works better.

Otherwise:

3. The value of  $\beta_1$  is calculated. according to the values of  $\beta_1$  and  $\beta_2$ , the following are obtained:
  - If  $\beta_2 < \beta < \beta_1$  then the aggregation approach works better.
  - If  $\beta = \beta_1 = \beta_2$  then both approaches are identical.
  - If  $\beta > \beta_1$  and  $\beta < \beta_2$  then the non-aggregation approach works better.

### Appendix L: Proof of theorem 5-3

This is a special case of Appendix K where  $\Delta$  defined at (K-2) is always positive, in this case when  $-1 < \theta < 1$  and  $-1 < \phi \leq 0$  the values of  $\beta_2$  is either smaller than zero or

greater than one ( $\beta_2 < 0$  or  $\beta_2 > 1$ ), Therefore we follow the same procedure as Appendix K and finally we get:

- If  $\beta < \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is greater than one and consequently the aggregation approach outperforms the non-aggregation one.
- If  $\beta = \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is equal to one and both approaches perform equally.
- If  $\beta > \beta_1$ , the ratio of  $MSE_{BA}/MSE_{AA}$  is smaller than one and the non-aggregation approach outperforms the aggregation one.

### Appendix M: Proof of theorem 7-3 for ARIMA(0,0,1) – Comparison at the aggregate level

In order to show that the aggregation approach is always outperforms non-aggregation one, we must show that the minimum value (lower bound) of the ratio ( $MSE_{BA}/MSE_{AA}$ ) is always greater than one, therefore to calculate the minimum value of  $MSE_{BA}/MSE_{AA}$ ,  $\theta$  and  $m$  should be equal to the smallest possible values of  $\theta=-1$  and  $m=2$ . By substituting these values in the  $MSE_{BA}/MSE_{AA}$ , we get

$$MSE_{BA}/MSE_{AA} = \left(1 + \frac{\alpha(1-\alpha)}{2}\right) \times \frac{2-\beta}{2-\alpha} \quad (M-1)$$

By considering  $0 < \alpha < 1$  and  $\beta < \alpha$  it's obvious that  $MSE_{BA}/MSE_{AA}$  is always greater than one.

When the smoothing constant values are very small, it is claimed that the ratio is equal to one. Therefore, we must show that  $\lim_{\alpha, \beta \rightarrow 0} (MSE_{BA}/MSE_{AA}) = 1$ , by considering (F-1) the following is given:

$$\lim_{\alpha, \beta \rightarrow 0} \left( 1 + \frac{\alpha(1-\alpha)}{2} \right) \times \frac{2-\beta}{2-\alpha} = 1 \quad (\text{M-2})$$

### Appendix N: Proof of theorem 9-3 for an ARIMA(1,0,0) - Comparison at aggregate level

In theorem 9-3 it is claimed that the aggregation approach is always works better than the non-aggregation approach. we must show that the minimum value (lower bound) of the ratio  $(MSE_{BA}/MSE_{AA})$  is always greater than one, so to calculate the minimum value of  $MSE_{BA}/MSE_{AA}$  with ARIMA(1,0,0) when  $-1 < \phi \leq 1/3$ ,  $\phi$  should be equal to  $\phi=.33$  and  $m=2$ . By substituting these values in the  $MSE_{BA}/MSE_{AA}$ , we get:

$$MSE_{BA}/MSE_{AA} = \left( \frac{2-\beta}{2-\alpha} \right) \times \left( \frac{0.8911+0.1089\beta}{.67+.33\alpha} \times \frac{3.5644+1.7822\alpha-1.7622\alpha^2}{4.741-.558\beta} \right) \quad (\text{N-1})$$

By considering  $0 < \alpha < 1$  and  $\beta < \alpha$  we can show that both parts of (N-1) are greater than one, therefore,  $MSE_{BA}/MSE_{AA}$  is always greater than one. In addition, when the smoothing constant is very small, the ratio is equal to one, we show that  $\lim_{\alpha, \beta \rightarrow 0} (MSE_{BA}/MSE_{AA}) = 1$ , now by considering (N-1) we have:

$$\lim_{\alpha, \beta \rightarrow 0} \frac{2-\beta}{2-\alpha} \times \left( \frac{0.8911+0.1089\beta}{.67+.33\alpha} \times \frac{3.5644+1.7822\alpha-1.7622\alpha^2}{4.741-.558\beta} \right) = 1 \quad (\text{N-2})$$

### Appendix O: Proof of theorem 9-3 for an ARIMA(1,0,0) - Comparison at aggregate level

By considering  $MSE_{BA}/MSE_{AA} < 1$  is equivalent to having the quadratic function (O-1) negative, which subsequently is equivalent to:

$$\begin{aligned} & (-\phi^m \Psi) \beta^2 + (2\phi^m \Psi - \Psi(1 - \phi^m) - 2\phi^m \eta + 2\xi) \beta \\ & + 2(1 - \phi^m)(\Psi - \eta) \end{aligned} \tag{O-1}$$

where

$$\Psi = \left( \begin{aligned} & \left( \frac{m}{1 - \phi^2} \right) + \left( \frac{\phi}{1 - \phi^2} \right) \left( \sum_{k=1}^{m-1} 2(m - k) \phi^{k-1} \right) + \frac{m^2 \alpha}{2 - \alpha} \left( \frac{1}{1 - \phi^2} \right) + \frac{2m^2 \alpha (1 - \alpha)}{(2 - \alpha)(1 - \phi + \alpha \phi)} \left( \frac{\phi}{1 - \phi^2} \right) \\ & - \frac{2m\alpha [(1 - \phi) + \phi(1 - \phi^{m-1})]}{(1 - \phi)(1 - \phi + \alpha \phi)} \left( \frac{\phi}{1 - \phi^2} \right) \end{aligned} \right)$$

$$\eta = \left( \frac{m}{1 - \phi^2} + \left( \sum_{k=1}^{m-1} 2(m - k) \phi^{k-1} \right) \left( \frac{\phi}{1 - \phi^2} \right) \right)$$

$$\xi = \left( \sum_{k=1}^m k \phi^{k-1} + \sum_{k=2}^m (k - 1) \phi^{2m-k} \right) \left( \frac{\phi}{1 - \phi^2} \right)$$

For the quadratic function given by (K-1), the value of the discriminant  $\Delta$  and the roots  $\beta_1$  and  $\beta_2$  can be defined as follows:

$$\Delta = (2\phi^m \Psi - \Psi(1 - \phi^m) - 2\eta \phi^m + 2\xi)^2 + 8\phi^m \Psi(1 - \phi^m)(\Psi - \eta) \tag{K-2}$$

$$\beta_1 = \frac{(2\phi^m \Psi - \Psi(1 - \phi^m) - 2\eta \phi^m + 2\xi) + \sqrt{\Delta}}{2\phi^m \Psi} \tag{K-3}$$

$$\beta_2 = \frac{-(2\phi^m \Psi - \Psi(1 - \phi^m) - 2\eta \phi^m + 2\xi) + \sqrt{\Delta}}{-2\phi^m \Psi} \tag{K-4}$$

We define the coefficient of  $\beta^2$  in (C-1) as follows

$$A = -\phi^m \Psi . \quad (\text{K-5})$$

If the discriminant  $\Delta < 0$ , there are no real roots for (K-1), therefore the sign of (K-1) is equivalent to the sign of  $A$ . We can show that when  $\Delta < 0$ ,  $A$  is always negative, consequently (K-1) is negative which means that  $MSE_{BA}/MSE_{AA}$  is smaller than one.

If  $\Delta > 0$ , (K-1) has two different roots  $\beta_1$  and  $\beta_2$ . By investigating the sign of  $\beta_1, \beta_2$  and  $A$ , we can determine the sign of (K-1) and consequently the performance superiority of each approach.

- If  $\beta_2 < 0$  and  $\beta_1 > 0$  then (K-1) is positive in the interval  $[\beta_2, \beta_1]$  and it is negative outside this interval.
- If  $\beta_2 > 1$ , we can show that  $0 < \beta_1 < \beta_2$ . (K-1) is negative in the interval  $[\beta_1, \beta_2]$  and it is positive outside this interval.
- If  $0 < \beta_2 < 1$ , we can show that  $\beta_1 > \beta_2$  then (K-1) is positive in the interval  $[\beta_2, \beta_1]$  and it is negative outside this interval.

By considering the above expressions, we get the following selection procedure.

1. The procedure is begun by calculating  $\Delta$  defined in (K-2), If  $\Delta < 0$  then the non-aggregation approach is always superior, otherwise the values of  $\beta_1$  and  $\beta_2$  defined in (K-3) and (K-4) are calculated.
2. If  $\beta_2 < 0$  or  $\beta_2 > 1$ , the value of  $\beta_1$  is calculated,
  - If  $\beta < \beta_1$  then the aggregation approach works better.
  - If  $\beta = \beta_1$  then both strategies are identical.
  - If  $\beta > \beta_1$  then the non-aggregation approach works better.

Otherwise:
3. The value of  $\beta_1$  is calculated. According to the values of  $\beta_1$  and  $\beta_2$  we have
  - If  $\beta_2 < \beta < \beta_1$  then the aggregation approach works better.

- If  $\beta = \beta_1 = \beta_2$  then both strategies are identical.
- If  $\beta > \beta_1$  and  $\beta < \beta_2$  then the non-aggregation approach works better.

### Appendix P: Covariance between demand $i,j$ and forecast $i,j$

The covariance between sub aggregate demand  $i$  and  $j$  defined as following:

$$\lambda_k = Cov(d_{i,t}, d_{j,t-k}) = \begin{cases} \frac{(1 - \phi_i \theta_j - \phi_j \theta_i + \theta_i \theta_j)}{1 - \phi_i \phi_j} (\rho_{ij} \sigma_i \sigma_j) & k = 0 \\ \frac{(\phi_i - \theta_i)(1 - \phi_i \theta_j)}{1 - \phi_i \phi_j} (\rho_{ij} \sigma_i \sigma_j) & |k| = 1, \\ \phi_i^{k-1} \lambda_1 & |k| > 1 \end{cases} \quad (P-1)$$

Similar to (P-1),  $\lambda'_k = Cov(d_{j,t}, d_{i,t-k})$  can be calculated where we substitute  $i$  by  $j$  and vice versa. By considering (A-1), the covariance between sub aggregate demand  $i$  and sub aggregate forecast  $j$  is calculated as follows:

$$\begin{aligned} Cov(d_{i,t}, f_{j,t}) &= Cov(d_{i,t}, \sum_{k=1}^{\infty} \alpha_j (1 - \alpha_j)^{k-1} d_{j,t-k}) = \alpha_j Cov(d_{i,t}, \sum_{k=1}^{\infty} (1 - \alpha_j)^{k-1} d_{j,t-k}) = \\ &\alpha_j (Cov(d_{i,t}, d_{j,t-1}) + (1 - \alpha_j) Cov(d_{i,t}, d_{j,t-2}) + (1 - \alpha_j)^2 Cov(d_{i,t}, d_{j,t-2}) + \dots) \quad (P-2) \\ &= \alpha_j \lambda_1 + \alpha_j (1 - \alpha_j) \phi_i \lambda_1 + \alpha_j (1 - \alpha_j)^2 \phi_i^2 \lambda_1 + \dots = \frac{\alpha_j \lambda_1}{1 - \phi_i + \alpha_j \phi_i}, \end{aligned}$$

Similar to (P-2), the covariance between sub aggregate demand  $j$  and sub aggregate forecast  $i$  is:

$$Cov(d_{j,t}, f_{i,t}) = \frac{\alpha_i \lambda'_1}{1 - \phi_j + \alpha_i \phi_j} \quad (P-3)$$

The covariance between subaggregate forecast  $i$  and  $j$  is as follows:

$$\begin{aligned}
 Cov(f_{i,t}, f_{j,t}) &= Cov(\alpha_i d_{i,t-1} + (1-\alpha_i) f_{i,t-1}, \alpha_j d_{j,t-1} + (1-\alpha_j) f_{j,t-1}) \\
 &= \alpha_i \alpha_j Cov(d_{i,t-1}, d_{j,t-1}) + \alpha_i (1-\alpha_j) Cov(d_{i,t-1}, f_{j,t-1}) \\
 &\quad + \alpha_j (1-\alpha_i) Cov(f_{i,t-1}, d_{j,t-1}) + (1-\alpha_i)(1-\alpha_j) Cov(f_{i,t-1}, f_{j,t-1}),
 \end{aligned}
 \tag{P-4}$$

Since  $Cov(f_{i,t}, f_{j,t}) = Cov(f_{i,t-1}, f_{j,t-1})$  and by considering  $\lambda'_k = Cov(d_{j,t}, d_{i,t-k})$  and substituting (P-1), (P-2), and (P-3) into (P-4), we get

$$Cov(f_{i,t}, f_{j,t}) = \left( \frac{1}{(1-(1-\alpha_i)(1-\alpha_j))} \right) \left( \alpha_i \alpha_j \lambda_0 + \frac{\alpha_i (1-\alpha_j) \alpha_j \lambda_1}{1-\phi_i + \alpha_j \phi_i} + \frac{\alpha_j (1-\alpha_i) \alpha_i \lambda'_1}{1-\phi_j + \alpha_i \phi_j} \right)
 \tag{P-5}$$

### Appendix Q: Proof of theorem 4-1

It's sufficient to show that the lower bound is greater than or equal to one, we use these facts that  $-1 \leq \rho_{12} \leq 1$ ,  $0.5 \leq p_1 + p_2 \leq 1$ , By considering the lower bounds of  $-1 \leq \rho_{12} \leq 1$ , the value of  $v_R$  is equal to zero, therefore we rewrite (4-41) as follows

$$\frac{V_{TD}}{V_{BU}} = \frac{(1-2\phi\theta + \theta^2)}{\left( \frac{2(1-\phi)((1+\theta^2 - 2\phi\theta) + \alpha\theta(1+\phi))}{(2-\alpha)(1-\phi + \alpha\phi)} \right)}
 \tag{Q-1}$$

Now by  $\phi=0.51$  and  $\theta=0$  for the interval of  $0.5 < \phi_1 = \phi_2 < 1$ ,  $-1 < \theta_1 = \theta_2 \leq 0$  we can calculate the upper bound of the ratio of  $V_{TD}/V_{BU}$ , now by substituting these values in (B-1) we have

$$\frac{V_{TD}}{V_{BU}} = \frac{(2-\alpha)(1-\phi + \alpha\phi)}{2(1-\phi)} = (1-0.5\alpha) \times (1+1.041\alpha)
 \tag{Q-2}$$

Now we can see that the minimum value of  $V_{TD}/V_{BU}$  is obtained when the smoothing constant  $\alpha$  becomes close to zero. Additionally, when  $\alpha=0$  the ratio equals to one.

### Appendix R: Proof of theorem 4-2, collary 4-1 and collary 4-2

By using these facts that  $-1 \leq \phi_1 = \phi_2 \leq 0.5$ ,  $-1 \leq \theta_1 = \theta_2 < 1$ ,  $-1 \leq \rho_{12} \leq 1$ ,  $0.5 \leq p_1 + p_2 \leq 1$ , the ratio of  $V_{TD}/V_{BU}$  for different values can be calculated from (4-41).

By considering the lower bounds of  $-1 \leq \rho_{12} \leq 1$ ,  $v_R$  in (4-41) is equal to zero, therefore (4-41) can be rewritten as

$$\frac{V_{TD}}{V_{BU}} = \frac{(1 - 2\phi\theta + \theta^2)}{\left( \frac{2(1 - \phi)((1 + \theta^2 - 2\phi\theta) + \alpha\theta(1 + \phi))}{(2 - \alpha)(1 - \phi + \alpha\phi)} \right)} \quad (\text{R-1})$$

To get the lower bound of (R-1) we need to set  $\phi$  to the minimum value  $\phi = -0.99$ , and  $\theta$  should be the maximum value,  $\theta = 0.99$  for the interval of  $-1 \leq \phi \leq 0.5$ ,  $-1 \leq \theta < 1$ , now we substitute these values in (R-1), so we have

$$\frac{V_{TD}}{V_{BU}} = \frac{3.9403(1.99 - 0.99\alpha)(2 - \alpha)}{3.98(3.9403 + 0.099\alpha)} \quad (\text{R-2})$$

Now by substituting  $\alpha = 0.01$  into (R-2) it is seen that  $V_{TD}/V_{BU} = 0.99$ .

To get the upper bound of  $MSE_{TD}/MSE_{BU}$ , the maximum values of  $0.5 \leq p_1 + p_2 \leq 1$  and  $-1 \leq \rho_{12} \leq 1$  are substituted into (4-41), we get

$$\frac{V_{TD}}{V_{BU}} = \frac{(1 - 2\phi\theta + \theta^2) + v_R}{\left( \frac{2(1 - \phi)((1 + \theta^2 - 2\phi\theta) + \alpha\theta(1 + \phi))}{(2 - \alpha)(1 - \phi + \alpha\phi)} \right)} \quad (\text{R-3})$$

where

$$v_R = \frac{2\left(\alpha_{TD}\left((1+\phi)(1-\theta)^2\right) - \alpha_{TD}^2\left(\phi + \phi\theta^2 - 2\theta\right) - \left(\alpha_{TD}(\phi - \theta)(1 - \phi\theta)(2 - \alpha_{TD})\right)\right)}{(2 - \alpha_{TD})(1 - \phi + \alpha_{TD}\phi)}$$

To get the lower bound of (R-3), the autoregressive parameters  $\phi$  is set to the maximum value  $\phi=0.5$ , and the moving average parameter  $\theta$  should be the minimum value,  $\theta=-0.99$ . Now, by substituting these values in (R-3), the following is given:

$$\frac{V_{TD}}{V_{BU}} = \frac{2.9701 + v_R}{\left(\frac{(2.9701 - 1.485\alpha)}{0.5(2 - \alpha)(1 + \alpha)}\right)} \quad (\text{R-4})$$

where

$$v_R = \frac{2\left[5.94\alpha_{TD} - 2.97\alpha_{TD}^2 - 2.23\alpha_{TD}(2 - \alpha_{TD})\right]}{0.5(2 - \alpha_{TD})(1 + \alpha_{TD})}$$

By substituting  $\alpha = \alpha_{TD} = 0.01$  into (R-4), the ratio equals to  $V_{TD}/V_{BU} = 1.01$ .

### Proof of Collary 5.1.

By substituting  $\alpha_{TD} = \alpha = 0.05, 0.15, 0.3$  in (R-2) and (R-4) the results presented in Table 2 can be obtained.

### Proof of Collary 5.2.

By substituting  $\phi = 0.51, \theta = 0.99$  and  $\phi = 0.99, \theta = 0.01$  into (R-1) and (R-3) the lower and upper bound of  $MSE_{TD}/MSE_{BU}$  can be obtained for the interval of  $0.5 < \phi_1 = \phi_2 < 1$ ,  $0 < \theta_1 = \theta_2 < 1$ . Finally, by substituting  $\alpha_{TD} = \alpha = 0.01, 0.05, 0.15, 0.3$  into that the results presented in Table 3 can be obtained.

## Résumé

---

Ce résumé a pour objectif de fournir une vision globale, les principaux objectifs et les étapes nécessaires de cette recherche. Nous commençons tout d'abord par définir certains termes clés dans le domaine de l'agrégation et de la prévision de la demande afin d'assurer une compréhension cohérente des concepts liés à ce travail de recherche. Par la suite, le contexte managérial et scientifique, l'aperçu et les objectifs de la recherche sont présentés. Enfin, une fois la démarche méthodologique adoptée dans ce travail est exposée, nous discutons les résultats et les contributions de ce travail.

### 1. Définitions

Une brève description des termes et des expressions clés utilisés dans ce travail de recherche est présentée dans les sections suivantes. Il s'agit des éléments appliqués tout au long de cette thèse.

- **Les séries chronologiques**

Makridakis et al (1998) définissent une série chronologique comme une séquence d'observations ordonnées dans le temps.

Bien que l'ordre soit généralement sur le temps, l'ordre peut également être considéré sur d'autres dimensions, comme l'espace (Harvey, 1993). Les séries chronologiques se produisent dans des domaines variés tels que l'agriculture, le commerce, l'économie, l'ingénierie, la géophysique, la médecine, les sciences sociales, etc. A titre d'illustration, dans le contexte de l'entreprise, le niveau de production annuel, la demande mensuelle de pièces détachées, le niveau des stocks hebdomadaires et des ventes quotidiennes sont toutes des séries chronologiques.

- **Séries chronologiques stationnaires**

Par série chronologique stationnaire, on entend une série dont les propriétés ne dépendent pas du temps durant lequel la série est observée (Makridakis et al., 1998).

Pour qu'un processus stochastique soit stationnaire, il faut que l'espérance mathématique de la série chronologique, la variance et l'auto-covariance de tout décalage d'ordre  $k$  soient constantes au cours du temps (Harvey, 1993). La classe la plus générale des modèles stationnaires pour la prévision des séries chronologiques est celui des processus autorégressifs et à moyenne mobile (ARMA).

#### ▪ **Séries chronologiques non-stationnaires**

La majorité des séries chronologiques existantes, en particulier dans les secteurs économiques et commerciaux sont non-stationnaires. Les séries chronologiques non-stationnaires peuvent se produire de plusieurs façons. Elles peuvent avoir des moyennes non constantes, des écarts et/ou autocovariances variant dans le temps, ou toutes ces propriétés simultanément. Les séries chronologiques concernant les tendances, saisonnalités et les séries cycliques sont des séries temporelles non-stationnaires (Wei, 2006). L'un des modèles typiques non-stationnaires est le processus autorégressifs et à moyenne mobile intégrée (ARIMA). Une série chronologique non-stationnaire peut être divisée en deux parties: i) séries chronologiques homogènes ii) séries temporelles non-homogène. Dans le premier cas, la moyenne est dépendante du temps. En calculant les différences entre les observations consécutives, une série chronologique homogène peut être convertie en série stationnaire: c'est la différenciation. Cependant, de nombreuses séries chronologiques non-stationnaires sont non-homogènes. La non-stationnarité de ces séries ne découle pas des moyennes dépendant du temps, mais résulte de la dépendance au temps de leurs variances et autocovariances.

#### ▪ **Méthodes de prévision**

Une méthode de prévision est une procédure pour estimer les observations futures. Elle dépend largement de la disponibilité des données. En cas d'indisponibilité, autrement dit si les données disponibles ne sont pas pertinentes pour les prévisions, les méthodes de prévision qualitatives doivent être utilisées. Il existe des approches structurées mieux développées pour l'obtention de bonnes prévisions sans l'aide de données historiques (Hyndman and Athanasopoulos, 2013). En revanche, les méthodes quantitatives peuvent être appliquées lorsque les conditions suivantes sont remplies:

1. Disponibilité des données numériques sur le passé,

2. Il est raisonnable de supposer que certains aspects des données passées vont se reproduire dans le futur.

Il existe un large éventail de méthodes de prévision quantitatives, souvent élaborées dans les disciplines spécifiques à des fins spécifiques. Chaque méthode a ses propres propriétés, sa précision et son coût qui doivent être considérés au moment de leur choix.

La plupart des méthodes de prévision quantitatives utilisent soit des séries chronologiques (collectées à des intervalles réguliers dans le temps) soit des données transversales (collectées à un moment précis). Les méthodes quantitatives de prévision sont divisées en deux catégories: 1) modèles de séries chronologiques ii) modèles explicatifs. Un modèle explicatif est très utile car il intègre des informations sur d'autres variables, plutôt que seulement les valeurs historiques de la variable à prévoir. Cependant, diverses raisons peuvent pousser un prévisionniste à sélectionner un modèle de série chronologique plutôt qu'un modèle explicatif. Premièrement, le système peut ne pas être compris, et même s'il l'était, il peut être extrêmement difficile de mesurer les relations qui déterminent son comportement. Deuxièmement, il est nécessaire de connaître ou de prévoir les diverses variables afin d'être en mesure d'anticiper sur la variable d'intérêt, et cela peut être très difficile. Troisièmement, la préoccupation principale peut être seulement de prévoir ce qui va se passer sans savoir pourquoi. En fin de compte, un modèle de séries chronologiques peut donner des prévisions plus précises qu'un modèle explicatif ou mixte (Hyndman and Athanasopoulos, 2013). Les modèles de séries chronologiques utilisés pour la prévision incluent des modèles ARIMA, le lissage exponentiel et les modèles structurels.

#### ▪ **Sélection de l'estimateur**

Afin d'évaluer l'impact de chaque approche d'agrégation sur la performance de la prévision, la sélection d'un estimateur dans un but d'extrapolation s'avère nécessaire. Dans cette étude, le lissage exponentiel simple (SES) est utilisé pour estimer la prévision de la demande. Il s'agit d'une méthode de prévision très populaire dans l'industrie car elle est intuitivement séduisante, facile à mettre à jour et possède des exigences minimales de stockage informatique des données. En outre, elle est optimale pour un processus non-

stationnaire à moyenne mobile intégrée, IMA(1) ou ARIMA (0,1,1). Bien que son application implique un comportement non-stationnaire de la demande, les valeurs suffisamment faibles de la constante de lissage (ou coefficient de lissage) introduisent des écarts mineurs de l'hypothèse de stationnarité, tandis que la méthode est aussi impartiale.

Le lissage exponentiel simple s'appuie sur des prévisions de la demande exponentiellement lissées. L'estimation est mise à jour à chaque période. Pour toute période de temps  $t$ , la procédure d'actualisation de la méthode SES est présentée suivant l'équation ci-après:

$$f_t = \alpha d_{t-1} + (1 - \alpha)f_{t-1} \quad (1)$$

où  $d_{t-1}$  est la demande à la période  $t-1$ ,  $f_t$  est la prévision à la période  $t$  et  $\alpha$  la constante de lissage.

Le coefficient  $\alpha$ , compris entre 0 et 1, s'applique à la dernière réalisation. Il s'agit de la constante de lissage choisie à ce niveau. Si  $\alpha$  est faible (par exemple, proche de zéro), plus de poids sera accordé aux observations plus loin dans le passé. Si par contre  $\alpha$  est grand (soit près d'un), plus de poids sera accordé aux observations plus récentes. Dans le cas extrême ( $\alpha = 1$ ), SES devient la méthode naïve. Dans ce travail de recherche, la méthode SES est préférée à la moyenne mobile (MA) et la méthode de prévision optimale, bien que ces méthodes de prévision peuvent être envisagées pour les futures recherches. Deux raisons justifient ce choix de méthode:

i) En moyenne, SES a tendance à donner de meilleures performances que la méthode MA, comme on l'observe dans une comparaison empirique de leur performance dans la compétition de prévision M3 (tel que rapportée par Makridakis and Hibon (2000)). De plus, SES correspond à un modèle intuitif séduisant contrairement à MA.

ii) En pratique, les décideurs ne veulent pas passer trop de temps et d'efforts pour examiner et définir les caractéristiques du processus de données avant de déterminer le modèle de prévision optimal, comme l'exige ARIMA. Par ailleurs, dans un cadre de planification de la production, les prévisions sont tenues sur une base périodique, parfois aussi souvent que quotidienne ou même horaire. Typiquement, la prévision est faite simultanément

pour plusieurs articles différents dans les systèmes informatiques avec un minimum d'intervention humaine. Par conséquent, il est relativement impossible de déterminer le modèle ARIMA optimal pour chaque élément à chaque mise à jour. Or, il serait utile de déterminer le montant du gain ou de perte en utilisant une méthode de prévision optimale au lieu de SES. Nous aborderons cette question dans les travaux futurs.

#### ▪ Indicateurs de précision

Un indicateur de précision est une mesure appliquée afin de juger l'efficacité du processus de prévision. Il existe de nombreux indicateurs permettant de mesurer la précision des prévisions. Dans le cadre de cette étude, la variance de l'erreur de prévision, aussi appelée erreur quadratique moyenne (MSE) est utilisée comme un indicateur de précision.

Le choix de sélectionner le MSE pour la comparaison théorique des méthodes considérées dans cette étude est justifié par le fait que ce dernier est une mesure de la précision mathématiquement attrayante. En outre, il se rapproche de la variance des erreurs de prévision (qui se compose de la variance des estimations produites par la méthode de prévision et la variance de la demande réelle), mais en diffère par le biais potentiel des estimations qui peut également être pris en compte. Étant donné que SES fournit des estimations non-biaisées des processus considérés dans ce travail, la variance des erreurs de prévision est égale à la MSE, *i.e.*  $MSE = Var(\text{erreur de la prévision})$

#### ▪ Agrégation de la demande

Un processus d'agrégation consiste à dériver le modèle de basse fréquence à partir du modèle à haute fréquence; cette dérivation peut être exercée dans le temps ou par l'intermédiaire des individus. L'agrégation dans le temps, aussi appelée agrégation temporelle, fait en particulier référence au processus par lequel une série de temps de basse fréquence (par exemple trimestrielle) est dérivée d'une série temporelle à haute fréquence (par exemple tous les mois) (Nikolopoulos et al., 2011). Comme montré dans les Figures 1 et 2, ce résultat est obtenu grâce à la somme de toutes les  $m$  périodes de données à haute fréquence, où  $m$  est le niveau d'agrégation.

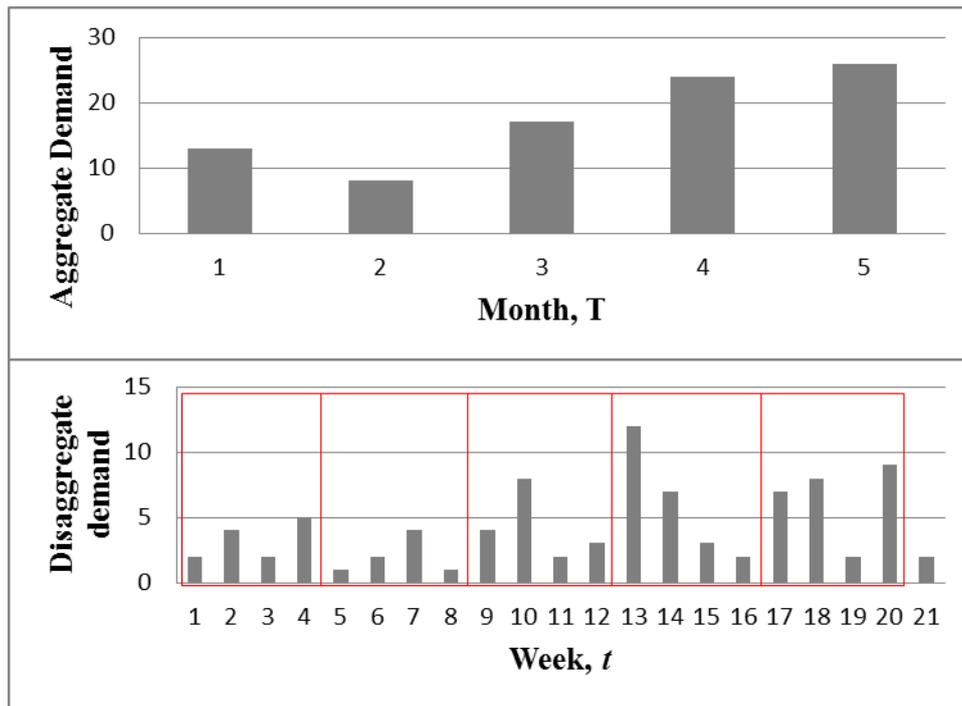


Figure 1: L'agrégation temporelle non-cumulée d'hebdomadaire à mensuelle

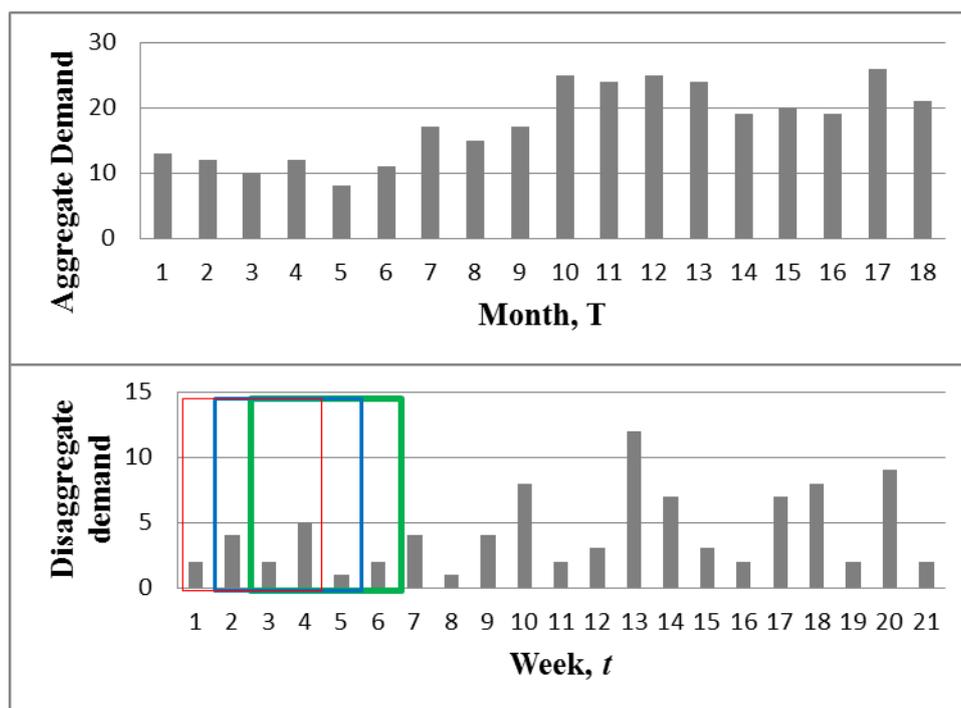
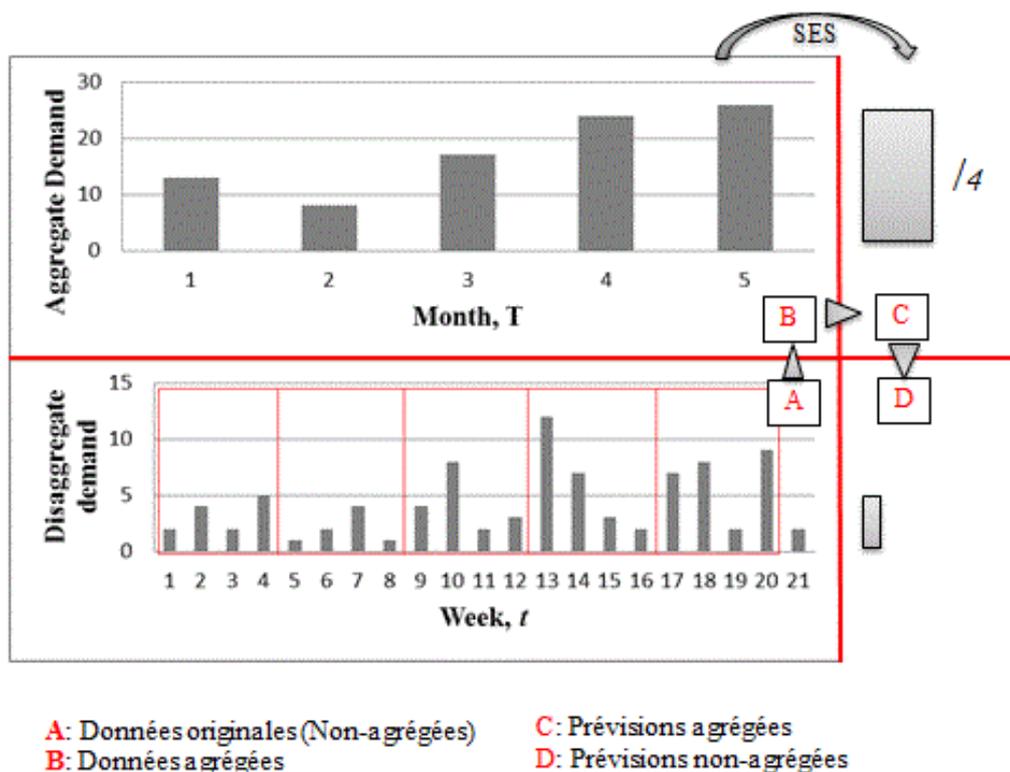


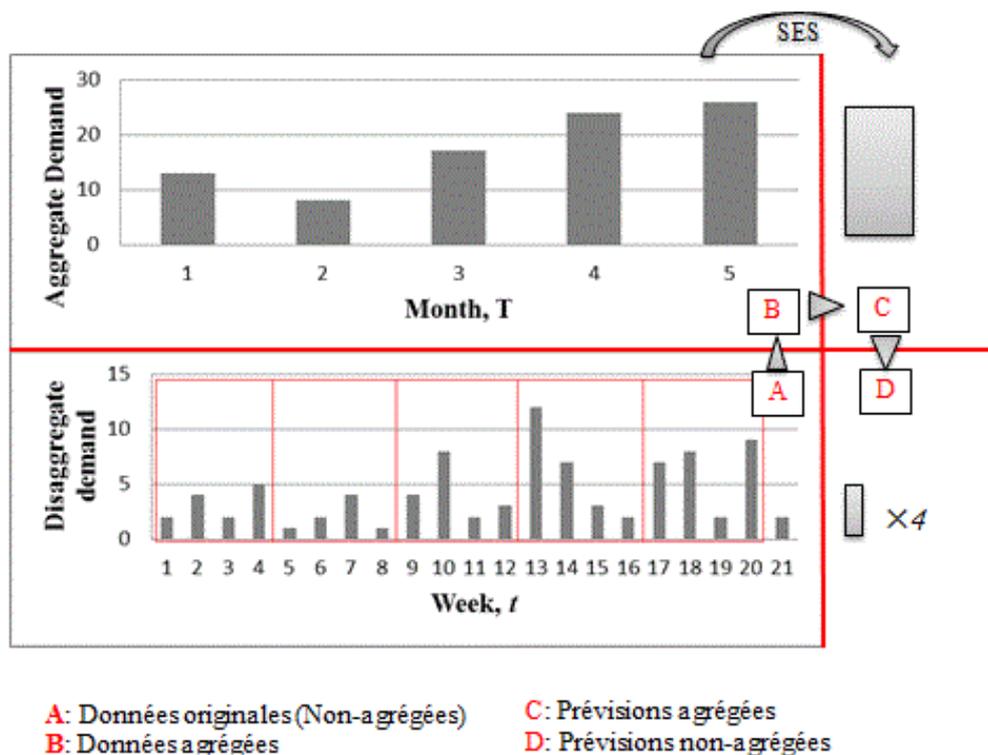
Figure 2: L'agrégation temporelle cumulée d'hebdomadaire à mensuelle

Il existe deux types différents d'agrégation temporelle: non-cumulée et cumulée. Dans le premier cas, les séries chronologiques sont divisées en segments consécutifs non-cumulée de temps, où la longueur de la tranche de temps est égale au niveau de l'agrégation. La demande agrégée est ainsi créée en additionnant les valeurs dans chaque tranche. Le nombre de périodes agrégées est  $[N/m]$ , où  $N$  est le nombre de périodes d'origine  $m$ , le niveau d'agrégation et  $[x]$  est la partie entière de  $x$ . En conséquence, le nombre de périodes de la demande agrégée est inférieur à la demande d'origine.

Souvent, pour avoir des prévisions comparables entre une approche d'agrégation et une approche de non-agrégation, si la comparaison est effectuée au niveau désagrégé, les prévisions agrégées doivent être désagrégées au niveau initial (en les divisant par le niveau d'agrégation). Par ailleurs, si la comparaison est effectuée au niveau agrégé, dans ce cas les prévisions initiales doivent être multipliées par le niveau d'agrégation. Ceci est illustré dans les **Figure 3** et **4** dans le cas de prévisions hebdomadaires et mensuelles.



**Figure 3:** Niveau de comparaison désagrégé

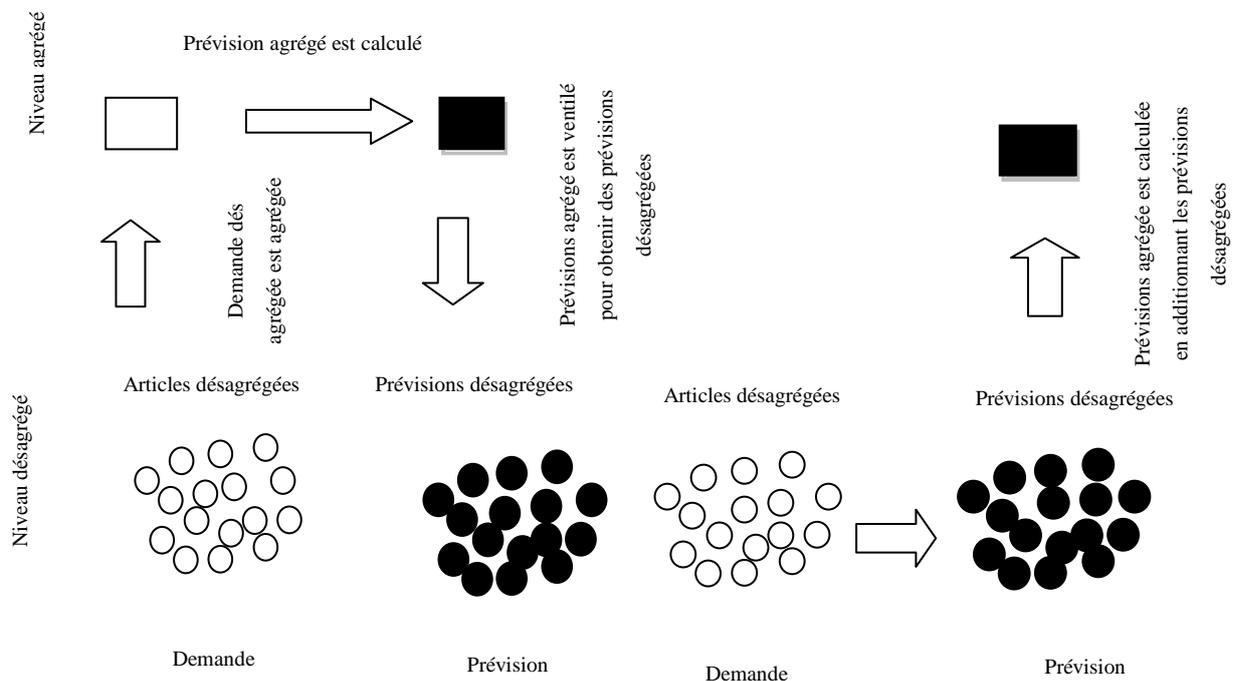


**Figure 4:** Niveau de comparaison agrégé

Un autre type d'agrégation (l'agrégation transversale encore appelée agrégation hiérarchique ou contemporaine) se fait au travers d'un certain nombre d'unités de gestion des stocks (SKU) à une période de temps précise afin de réduire la variabilité (Silvestrini and Veredas, 2008). Les approches existantes de prévision transversale impliquent généralement soit une approche ascendante (BU), soit une approche descendante (TD), voire une combinaison des deux. Lorsque la prévision au niveau agrégé est en question, cette dernière implique l'agrégation des prévisions des unités de gestion des stocks individuelles au niveau du groupe, tandis que la deuxième concerne la prévision directement au niveau du groupe (i.e. ceci exige premièrement l'agrégation de la demande, puis extrapoler directement la prévision au niveau global). Lorsque l'accent est mis sur la prévision au niveau désagrégé, l'approche BU concerne l'extrapolation directe au niveau désagrégé alors que TD implique la désagrégation des prévisions agrégés produites directement au niveau du groupe.

Comme l'illustre la **Figure 5**, L'approche TD se compose des étapes suivantes: i) les demandes sous-agrégats sont agrégées; ii) production des prévisions de demande agrégée via

la méthode SES au niveau agrégé, et iii) la prévision est désagrégée pour revenir à son niveau initial en appliquant une méthode de désagrégation appropriée, si une prévision désagrégée est exigée. Dans l'approche BU: i) les prévisions de la demande désagrégée sont produites directement pour les articles désagrégés; ii) la prévision agrégée est obtenue en combinant les prévisions individuelles pour chaque SKU, soit potentiellement un modèle de prévision séparé utilisé pour chaque élément de la famille de produits (Zotteri et al., 2005). Ces approches sont présentées schématiquement dans la **Figure 5**. Nous adoptons ainsi le style de présentation de Mohammadipour et al. (2012).



**Figure 5:** Schéma de TD (gauche) et BU approches (droite)

## 2. Contexte Managérial

La prévision de la demande est le point de départ de la plupart des activités de la planification et du contrôle des organisations. En outre, l'un des défis les plus importants des

sociétés modernes est l'incertitude de la demande (Chen and Blue, 2010). L'existence d'une forte variabilité de la demande des articles à grande ou à faible rotation pose des difficultés considérables en termes de prévision et de gestion de stock (Chen et al., 2000; Syntetos and Boylan, 2005; Wemmerlov and Whybark, 1984).

Il existe plusieurs approches qui peuvent être utilisées pour réduire l'incertitude de la demande et par conséquent améliorer la performance de la prévision (et la gestion des stocks) d'une entreprise. Une approche intuitivement attrayante, connue pour être efficace, est l'agrégation de la demande (Chen et al., 2007). Une possibilité est l'agrégation temporelle. Une autre approche d'agrégation souvent appliquée dans la pratique est l'agrégation transversale, c'est-à-dire l'agrégation des données de plusieurs SKUs. Cette approche est équivalente aussi à l'agrégation des données d'un seul SKU à travers d'un certain nombre de dépôts ou des lieux d'stockage. Naturelles et utiles dans la pratique, des formes d'agrégation associées impliquent également la consolidation géographique des données ou le regroupement entre les marchés. Bien qu'il n'y ait pas d'étude empirique qui documente la mesure dans laquelle l'agrégation a lieu dans un contexte pratique, il s'agit d'une approche qui est connue pour être efficace parmi les professionnels en raison de son attrait intuitif. En termes pratiques, la prestation dépend du type d'agrégation et bien sûr des caractéristiques des données. Une agrégation transversale par exemple conduit généralement à la réduction de la variance. Cela est dû au fait que les fluctuations dans les données d'une série chronologique peuvent être compensées par les fluctuations présentes dans une autre série (Widiarta et al., 2009). Contrairement à l'agrégation transversale, dans l'agrégation temporelle la variance augmente. Cependant, il peut facilement être montré que l'agrégation temporelle peut réduire le coefficient de variation de la demande. Dans tous les cas, l'avantage implicite associé à la facilité de mise en œuvre de ces approches les rend un choix populaire dans l'industrie.

En pratique, la demande peut être classée comme intermittente ou à forte rotation. Dans le premier cas, l'agrégation temporelle de la demande entraînerait la réduction de la présence d'observations nulles et, plus généralement, la réduction des incertitudes dans le second cas. Les articles à demande intermittente (comme pièces de rechange) sont connus pour causer des difficultés considérables en termes de prévision et modélisation des stocks. La présence de zéros a des implications importantes en raison des trois raisons suivantes. Tout d'abord, la difficulté à capturer les caractéristiques des séries chronologiques étudiées et des

modèles de prévision standards qui leurs correspondent. Deuxièmement, la difficulté de s'adapter à une distribution statistique standard telle que la loi normale. Troisièmement, les écarts par rapport aux hypothèses de modélisation de stock standard et leurs formulations. Ceux-ci rendent la gestion de ces éléments un exercice très difficile. L'agrégation temporelle est connue pour être largement appliquée dans les milieux militaires (données très rares), le secteur après-vente (pièces détachées ou de service), etc. Des études empiriques récentes (Babai et al., 2012; Nikolopoulos et al., 2011) dans ce domaine ont abouti à des résultats très prometteurs en soulignant également la nécessité d'une analyse plus théorique. Bien que le domaine de la prévision à l'aide de l'agrégation temporelle dans un contexte de demandes intermittentes est très intéressant tant d'un point de vue académique et professionnel, dans cette recherche le contexte des demandes à forte rotation, qui reste le contexte le plus rencontré, est celui pris en compte. L'analyse dans un contexte de demandes intermittentes est une voie intéressante de recherches futures et cette question est abordée avec plus en détail dans le dernier chapitre de cette thèse.

En plus de la réduction de l'incertitude de la demande associée à l'approche de l'agrégation temporelle discutée ci-dessus, il y a une question importante dans un processus de prévision où l'agrégation temporelle peut être utile. Il est appelé "horizon de la prévision" qui détermine la limite de la prévision future. En règle générale, plus on regarde loin dans le futur, plus la précision décroît. C'est aussi l'un des domaines où l'agrégation temporelle peut améliorer la précision des prévisions, parce que comme nous regardons plus loin dans l'avenir, la vision à long terme devient plus importante et la méthode d'agrégation temporelle peut utiliser cette information mieux que les approches classiques. Donc, l'approche d'agrégation temporelle peut aussi être très efficace lorsque les professionnels ont besoin de prévisions à long terme au lieu d'une prévision pour une seule période future. D'un point de vue théorique, l'accent à ce jour a été principalement sur l'agrégation transversale. En outre, la plupart des logiciels de prévision prend en charge l'agrégation des données, ce serait aussi couvrir seulement l'agrégation transversale. La considération de l'agrégation temporelle a été quelque peu négligée par les éditeurs de logiciels et les chercheurs malgré la possibilité d'ajouter plus de valeur en pratique. Dans ce travail, l'objectif est de faire progresser l'état actuel des connaissances dans le domaine de la prévision de la demande à l'aide de l'agrégation temporelle.

Dans les discussions ci-dessus, l'effet de l'agrégation temporelle sur un seul SKU est considéré. Alors qu'en réalité, il y a souvent de nombreuses séries chronologiques qui peuvent être organisées de façon hiérarchique et groupées à différents niveaux dans les groupes basés sur des références de produits, des clients, de la géographie ou d'autres caractéristiques (Hyndman et al., 2011). Le niveau hiérarchique auquel la prévision est effectuée dépend du besoin de chaque fonction. En ce qui concerne les produits (ou références), en particulier, la prévision au niveau SKU individuel est nécessaire pour la gestion des stocks, les prévisions de la famille de produits peuvent être requises pour le programme directeur de production. Les prévisions à travers d'un groupe d'articles commandés auprès du même fournisseur peuvent être nécessaires dans le but de regrouper les commandes. Les prévisions à travers des articles vendus à un grand client spécifique peuvent impacter le transport, les décisions de routage, etc. Une approche a priori intéressante pour obtenir des prévisions de niveau supérieur est l'agrégation transversale, ce qui implique généralement soit une approche TD ou une approche BU (ou une combinaison des deux). Une question importante qui a attiré l'attention de nombreux chercheurs et professionnels au cours de ces dernières décennies est l'efficacité de ces approches de prévision transversales.

Les approches de prévision BU et TD sont extrêmement utiles pour améliorer la précision des prévisions et des plans au sein d'un processus S&OP (la planification des ventes et des opérations) (Lapide, 2006). Le S&OP est un processus multifonctionnel qui implique les gestionnaires de tous les départements (ventes, service client, chaîne logistique, marketing, fabrication, achats et finances), où chaque département a besoin de différents niveaux des prévisions de la demande (Lapide, 2004). Par exemple, dans le marketing (Dekimpe and Hanssens, 2000), la prévision du chiffre d'affaires par groupes de produits et par marques est nécessaire. Les services commerciaux traitent avec des prévisions de ventes par les comptes clients et/ou des canaux de vente. Les gestionnaires de la chaîne d'approvisionnement demandent les prévisions au niveau du SKU, tandis que la finance a besoin de prévisions qui sont agrégées dans les unités budgétaires en termes de revenus et de coûts (Bozos and Nikolopoulos, 2011). Afin de produire les prévisions requises, la demande et/ou les prévisions devraient être agrégés et/ou désagrégés à différents niveaux. Il s'agit de l'application des approches TD et BU ou une combinaison des deux (Lapide, 2004, 2006).

### 3. Contexte scientifique

L'agrégation a été largement discutée dans la littérature académique depuis les années 1950 (Quenouille, 1958). Elle est considérée comme un moyen de réduire les fluctuations de la demande et le degré d'incertitude. Il a été démontré par Theil (1954), Yehuda and Zvi (1960), et Aigner and Goldfeld (1974) que l'incertitude de la demande peut être efficacement réduite par l'agrégation et une bonne prévision de la demande. Dans la littérature de la planification de la chaîne logistique et la planification de la demande, l'agrégation de la demande est connue comme un approche de mutualisation des risques pour réduire les fluctuations de la demande afin d'avoir une planification des matières/capacité plus efficace (Chen and Blue, 2010). Dans le domaine de l'agrégation temporelle, il y a à la fois des études théoriques et empiriques discutées dans la littérature. Cependant, la plupart de ces études sont dans le domaine de l'économie. Les propriétés du processus agrégé sont fournies sur la base des données non-agrégées. De plus, l'effet de l'agrégation temporelle sur la prévision est évalué par l'ajustement d'un modèle et de l'estimation des paramètres. Amemiya and Wu (1972) ont évalué l'effet de l'agrégation temporelle non-cumulée lorsque la série originale suit un processus autorégressif d'ordre  $p$ , AR ( $p$ ). En considérant le ratio des MSE de la prévision non-agrégée et agrégée (3 prédicteurs linéaires ont été considérés au niveau agrégé, ils ont montré que l'approche d'agrégation performe mieux que l'approche non-agrégée. Tiao (1972) a étudié l'effet de l'agrégation temporelle non-cumulée sur un processus non-stationnaire moyenne mobile intégrée d'ordre  $(p,q)$ , l'IMA  $(p,q)$ . Une espérance conditionnelle est appliquée pour obtenir une prévision à l'horizon d'une période au niveau agrégé basé sur les séries non-agrégées et agrégées. Par la suite, l'efficacité des prévisions agrégées a été définie comme le ratio de la variance de l'erreur de prévision de la série non-agrégée à la série agrégée lorsque le niveau d'agrégation est grand. On montre que lorsque  $d = 0$  et le niveau d'agrégation est très grand, alors le ratio en question est égal à un et l'avantage comparatif de l'utilisation des prévisions non-agrégées augmente avec  $d$ .

Peu d'études récentes ont évalué l'effet de l'agrégation temporelle sur la prévision et la gestion des stocks par des recherches empiriques. Nikolopoulos et al. (2011) ont empiriquement analysé les effets de l'agrégation temporelle sur la prévision de demandes intermittentes et ils ont proposé la méthodologie ADIDA. Il est démontré que la méthodologie ADIDA peut en effet apporter des améliorations considérables en termes de précision des

prévisions. Enfin, Babai et al. (2012) ont également étendu l'étude décrite ci-dessus (Nikolopoulos et al., 2011) afin d'examiner les implications de la méthodologie ADIDA sur les stocks en considérant une politique à suivi périodique appelée politique avec niveau de recombplètement (ordre-up-to-level). Les chercheurs ont conclu qu'une technique simple comme l'agrégation temporelle peut être aussi efficace que les approches mathématiques complexes de prévision des demandes intermittentes.

Au meilleur de notre connaissance, les seules études directement pertinentes pour notre travail sont celles par Amemiya and Wu (1972) et Tiao (1972) pour les processus AR et MA respectivement. Ces travaux ont porté sur la caractérisation de la série de la demande agrégée en plus de l'évaluation de la performance des prévisions. Cependant, les résultats présentés dans ces travaux restent préliminaires alors que le contexte expérimental peut également être critiqué en termes des procédures d'estimation considérées. De plus, aucun résultat empirique n'a été fourni. Par conséquent, l'absence des conditions qui déterminent la supériorité d'une approche, en matière de prévision de la demande, est évidente. Il n'est pas clair si l'approche d'agrégation fournit des prévisions plus précises que celle de la non-agrégation, et vice versa. Par conséquent, la motivation derrière cette partie de l'étude est l'absence de l'analyse théorique en ce qui concerne l'effet de l'agrégation temporelle sur la prévision de la demande. Dans cette recherche, l'évaluation analytique est appliquée pour déterminer les conditions de supériorité de chaque approche. La recherche est commencée avec le processus simple ARMA d'ordre un. Cependant, l'analyse peut être effectuée pour les processus d'ordres supérieurs mais les résultats deviennent plus complexes à présenter donc ceci est considéré dans les recherches futures.

Dans le domaine de l'agrégation transversale, la plupart de la littérature de la prévision s'est penchée sur les performances comparées des approches TD et BU. Les conclusions en ce qui concerne les performances de ces approches sont mélangées.

Certains auteurs comme Theil (1954), Grunfeld and Griliches (1960), Schwarzkopf et al. (1988), et Narasimhan et al., 1985(1985) ont fait valoir que l'approche TD performe mieux que BU, d'autre part, des auteurs comme Orcutt et al. (1968), Edwards and Orcutt (1969), Dunn et al. (1976), Dangerfield and Morris(1988) and Gross and Sohl (1990) ont constaté que l'approche BU est performante et enfin quelques autres auteurs comme Barnea and Lakonishok (1980), Flidner (1999) and Widiarta et al.(2007, 2008, 2009) adoptent une

approche contingente et analysent les conditions dans lesquelles une méthode produit des prévisions plus précises que les autres.

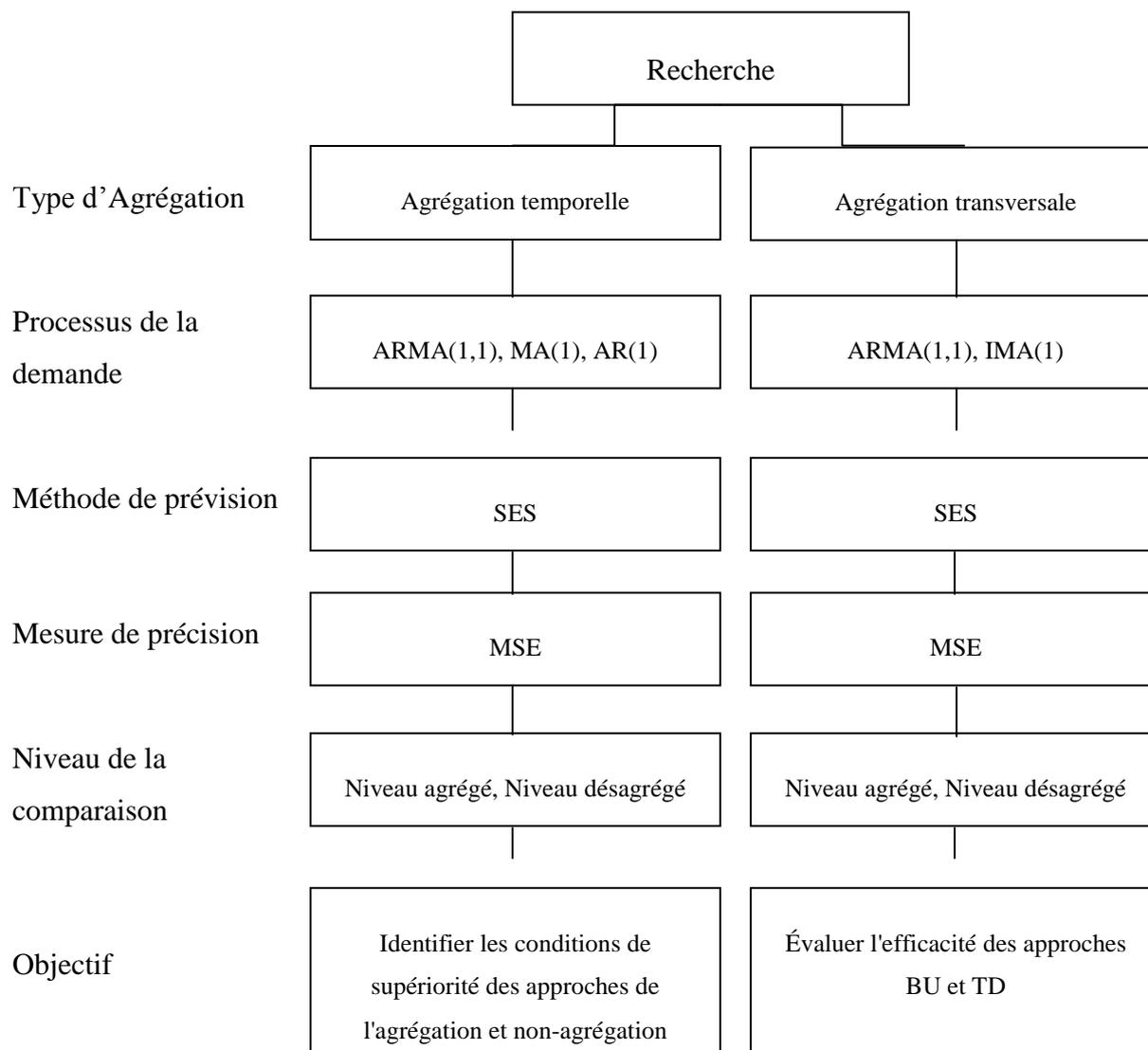
Dans cette thèse, l'efficacité de la BU et la TD est évaluée. Les travaux présentés par Widiarta et al. sont étendus dans cette thèse en considérant un processus de demande stationnaire plus général ARIMA(1,0,1) et un processus non-stationnaire ARIMA(0,1,1). Par ailleurs, la comparaison est effectuée tant au niveau désagrégé et agrégé. En outre, la supériorité de chaque approche est examinée en utilisant un ensemble de données réelles.

#### **4. Aperçu de la recherche**

L'agrégation est un moyen efficace pour réduire la variabilité de la demande. De plus, il permet aux prévisionnistes d'obtenir différents niveaux de prévisions dans le temps et des niveaux hiérarchiques. Selon le niveau des prévisions, nous produisons d'abord les prévisions et les agrégeons par la suite soit nous regroupons d'abord les séries originales individuelles pour obtenir la demande agrégée et puis de produire la prévision agrégée. Dans ce dernier cas, un mécanisme de désagrégation est nécessaire pour obtenir les prévisions désagrégées. Dans cette recherche, l'impact de l'agrégation sur la prévision de la demande est évalué. Pour montrer l'effet de l'agrégation sur la prévision de la demande, deux types d'agrégation sont considérés: *i)* l'agrégation temporelle et *ii)* l'agrégation transversale. Notre aperçu de la recherche est résumé dans la Figure 6.

On suppose que la série chronologique suit un processus de type ARIMA et la méthode de prévision est SES.

Dans l'agrégation temporelle, il est supposé que la demande désagrégée suit un processus stationnaire autoregressif moyenne mobile d'ordre un, ARIMA (1,0,1), ce qui veut dire que leurs cas particuliers, moyenne mobile d'ordre un, ARIMA (0,0,1) et l'autoregressif d'ordre un, ARIMA (1,0,0) sont également considérés. Ensuite, il est discuté si des données désagrégées ou des données agrégées doivent être utilisées pour fournir les prévisions requises. De plus, les conditions dans lesquelles, une approche performe mieux que l'autre sont présentées.



**Figure 6:** Vue d'ensemble de la recherche

Dans l'agrégation transversale, l'efficacité des approches BU et TD, pour fournir des prévisions désagrégées et agrégées, est analysée. On suppose que la série désagrégée suit soit un processus stationnaire autoregressif moyenne mobile d'ordre un, ARIMA(1,0,1) soit un processus non-stationnaire moyenne mobile intégrée d'ordre un, IMA (1,1). Dans cette partie de la thèse, la variance de l'erreur de prévision est utilisée pour comparer la performance de chaque approche. La variance de l'erreur de prévision est équivalente à la MSE en considérant une méthode de prévision non biaisée. Les variances des erreurs de prévision sont obtenues sur la base de la demande désagrégée et agrégée. Les comparaisons sont effectuées au niveau de la demande désagrégée et agrégée. Les conditions dans lesquelles chaque approche surpasse les autres sont mathématiquement identifiées. L'analyse mathématique est complétée

par une étude numérique pour valider les résultats théoriques. De plus, l'étude numérique est appliquée pour évaluer en détail les conditions de supériorité de l'approche en relaxant certaines hypothèses considérées dans l'évaluation analytique. Ensuite, les résultats sont validés empiriquement (à l'aide des simulations sur un ensemble de données réelles fournies par un hypermarché européen). Enfin, des leçons managériales très importantes sont dérivées et des suggestions concrètes sont proposées aux professionnels qui s'intéressent aux problèmes de prévision et de gestion des stock.

Dans cette recherche, l'ordre des processus de type ARIMA est limité à un car ce type de processus est plus observé dans la littérature pour les séries non-saisonnnières, outre l'objectif principal qui est de tirer plusieurs éclairages clés pour les managers. Par conséquent, nous allons limiter notre attention aux processus AR(1), MA(1) et ARMA(1,1). Toutefois, il convient de noter que l'extension du travail à analyser des cas plus généraux tels que AR ( $p$ ), MA ( $q$ ), voire ARMA ( $p, q$ ) est faisable, mais l'analyse et la présentation des résultats deviendraient complexe. Cette analyse sera examinée dans les travaux futurs.

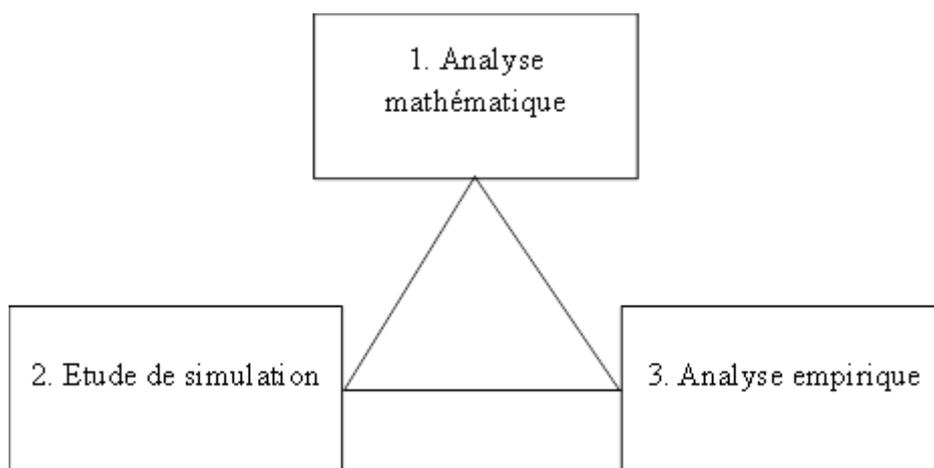
L'objectif principal de cette recherche est d'analyser les effets de l'agrégation sur la prévision de la demande. Cet effet est examiné par l'analyse mathématique et l'étude de simulation. L'analyse est complétée en examinant les résultats sur un ensemble de données réelles. Basé sur le contexte scientifique et managérial de la recherche et des motivations, six objectifs ont été formulés pour cette recherche:

7. Evaluer analytiquement l'effet de l'agrégation temporelle non-cumulée sur la prévision lorsque la série de base suit un processus stationnaire de type ARMA.
8. Identifier les conditions dans lesquelles l'approche d'agrégation temporelle performe mieux que celle de non- agrégation, et vice versa.
9. Déterminer le niveau d'agrégation optimale qui maximise les avantages de l'approche d'agrégation temporelle.
10. Examiner l'efficacité des approches BU et TD afin de prévoir la demande désagrégée et agrégée dans un environnement stationnaire et non-stationnaire.
11. Analyser l'effet des paramètres du processus et de contrôle sur la supériorité de l'approche dans les agrégations temporelles et transversale.

12. Tester la validité empirique et l'utilité des résultats théoriques et de simulation sur un large ensemble de données réelles.

## 5. Methodologie

La recherche suit trois méthodes de recherche, l'analyse mathématique, la simulation et l'étude empirique. La relation entre les trois méthodes est illustrée dans la Figure 7.



**Figure 7:** Méthodologie

Premièrement, l'analyse mathématique est appliquée afin d'examiner la supériorité de l'approche d'agrégation et de dévoiler les conditions dans lesquelles cette approche donne des résultats plus précis par rapport à l'approche classique. La variance théorique de l'erreur de prévision associée à chaque approche est calculée pour tous les processus de la demande à l'étude. Ceci est mené afin d'identifier les conditions de la supériorité de chaque approche. L'étude de simulation est utilisée pour les raisons suivantes:

- Pour tester et valider les résultats de l'analyse théorique.
- Pour relaxer les hypothèses prises en compte dans l'évaluation mathématique.

Enfin, les résultats de cette thèse sont testés sur des données empiriques réelles pour évaluer la validité et l'applicabilité pratique des principaux résultats de l'étude. Par

conséquent, l'analyse empirique nous aiderait à tester l'applicabilité des résultats dans des situations réelles.

## 6. Résultats et contributions

Dans la première partie de cette étude, l'impact de l'agrégation temporelle sur la prévision de la demande a été évalué par l'analyse théorique, la simulation et l'investigation empirique. Les évaluations sont basées sur le calcul de l'erreur quadratique moyenne (MSE) avant et après l'agrégation ( $MSE_{BA} / MSE_{AA}$ ) et les comparaisons sont menées à la fois au niveau désagrégé et agrégé. Il est supposé que la demande suit un processus stationnaire ARIMA(1,0,1), ARIMA(0,0,1) et ARIMA(1,0,0) et un lissage exponentiel simple est utilisé comme la méthode de prévision. Les conditions dans lesquelles l'approche d'agrégation performe mieux que celle de non-agrégation sont identifiées.

Les résultats de cette recherche concernant l'agrégation temporelle sont les suivantes:

- Les conditions de la supériorité des approches de l'agrégation et la non-agrégation sont identifiées. Les valeurs des points de rupture sont déterminées pour des valeurs données du niveau d'agrégation, et la constante de lissage associée à la série de la demande initiale. Il en résulte des règles théoriques montrant la performance de chaque approche aux niveaux de la comparaison désagrégée et agrégée.
- La performance de l'approche d'agrégation se trouve généralement améliorée quand le niveau d'agrégation augmente. Le taux d'amélioration cependant, est plus faible pour les processus ARIMA (1,0,1) et ARIMA (1,0,0) par rapport à celui de ARIMA(0,0,1). Dans tous les processus, le niveau d'agrégation optimal est la valeur la plus élevée sur un intervalle donné du niveau d'agrégation.
- La performance de l'approche d'agrégation s'améliore lorsque la valeur de la constante de lissage employée à la série agrégée baisse. Les résultats de l'analyse montrent que lorsque le niveau d'agrégation augmente, l'auto-corrélation de la série est réduite, ce qui nécessite l'emploi de valeurs faibles des constantes de lissage.

- En général, on constate que pour des valeurs d'auto-corrélation positivement élevées dans la série originale, la méthode d'agrégation peut être dépassée par celle de non-agrégation.
  - lorsque l'on compare au niveau désagrégé et où l'auto-corrélation est extrêmement positive, (c.à.d hautes valeurs positives de  $\phi$  dans le processus ARIMA (1,0,0) ou des valeurs négatives élevées de  $\theta$  et de valeurs positives élevées de  $\phi$  dans le processus ARIMA (1,0,1)), aucun niveau d'agrégation ne permet d'améliorer la précision des prévisions. Par conséquent, l'approche non-agrégation fournit toujours des prévisions plus précises. C'est un résultat intuitif car, à tout moment de la période de demande les informations les plus récentes sont précieuses. Dans un tel cas, l'approche désagrégée fonctionne mieux car elle exploite pleinement ces informations récentes.
  - Toutefois, lorsque la comparaison est effectuée au niveau agrégé, même pour des valeurs extrêmement positives de l'auto-corrélation, l'approche d'agrégation peut performer mieux que celle de non-agrégation en fonction du niveau d'agrégation. Pour les valeurs faibles du niveau d'agrégation, l'approche non-agrégation fonctionne mieux. Néanmoins, en augmentant le niveau d'agrégation, l'approche agrégation surpasse la non-agrégation. C'est parce que la comparaison est effectuée au niveau agrégé où une prévision cumulative sur l'horizon de  $m$  périodes est nécessaire. Comme le niveau d'agrégation et par conséquent l'horizon de prévision augmente, la précision des prévisions résultant de l'approche non-agrégation se détériore et cède à une supériorité en faveur de l'approche d'agrégation.
- Pour les valeurs négatives d'auto-corrélation ou les valeurs positives faibles, l'approche d'agrégation est préférable quel que soit le niveau de comparaison. Lorsque l'auto-corrélation de la demande est négative ou positive faible, l'information récente n'est pas cruciale, puis une vue à plus long terme sur la demande est préférable. Ceci peut être obtenu comme indiqué en sélectionnant un niveau d'agrégation élevé et une valeur faible de constante de lissage.

Dans la deuxième partie de cette recherche, l'efficacité des approches BU et TD est analytiquement évaluée pour prévoir la demande au niveau agrégé et désagrégé quand la série désagrégée suit soit un processus moyenne mobile intégrée d'ordre un, ARIMA (0,1,1), soit un processus autoregressif moyenne mobile d'ordre un, ARIMA (1,0,1). La méthode de prévision appliquée est une procédure de lissage exponentiel simple (SES). Les résultats des analyses théoriques ont été complétés par une étude de simulation à la fois au niveau agrégé et désagrégé ainsi que l'expérimentation avec un ensemble des données empiriques relatives à un hypermarché européen. Les développements sont basés sur la détermination de la variance de l'erreur de prévision pour les approches TD et BU. Les comparaisons sont menées tant au niveau désagrégé et agrégé.

Les résultats de cette recherche concernant l'agrégation transversale sont les suivantes:

- Lorsque les paramètres de processus de tous les articles désagrégés sont identiques, il n'y a pas de différence significative entre les approches TD et BU en prévision du niveau agrégé tant que la constante de lissage optimale est celle utilisée pour les deux approches. En outre, la performance des approches TD et BU est identique lorsque les constantes de lissage utilisées pour tous les articles désagrégés et la demande agrégée sont identiques.
- Lorsque l'auto-corrélation des articles désagrégés est très positive, les approches BU et TD affichent la même performance indépendamment des valeurs de corrélation croisée.
- Pour fournir les prévisions agrégées, l'approche TD performe mieux que BU lorsque les corrélations croisées entre les articles désagrégés sont (très) positives, la d'auto-corrélation d'un article est positivement élevée et celle d'un autre est négative ou faible positive.
- BU peut performer mieux que TD pour fournir les prévisions agrégées lorsque les articles désagrégés suivent différents modèles de fluctuation (corrélation croisée négative). La TD ne semble pas être très précise quand les articles désagrégés ont des profils de demande différents.
- Pour fournir des prévisions désagrégées, l'approche BU performe mieux que TD lorsque l'auto-corrélation d'au moins un élément de la famille est positive et la

constante de lissage est fixée à sa valeur optimale pour les deux approches, ceci est indépendamment de la corrélation croisée, les poids de la méthode de désagrégation, et des valeurs des paramètres du processus. Le degré de supériorité de l'approche BU pour les processus non-stationnaires est beaucoup plus élevé par rapport à celui des processus stationnaires lorsque l'on compare au niveau désagrégé.

- On constate que pour l'auto-corrélation négative ou positive faible, les approches BU et TD montrent presque la même performance pour prévoir la demande désagrégée lorsque les constantes de lissage optimales sont utilisées. En outre, la différence entre ces deux approches peut aller jusqu'à 1%.
- La performance de BU est généralement améliorée par la diminution de la corrélation croisée, passant des valeurs positifs à négatifs. La performance de l'approche TD se détériore par la baisse de la corrélation croisée. Pour les valeurs de corrélation croisée très négatives, BU est toujours préférée. C'est généralement le cas pour la comparaison au niveau agrégé et désagrégé.
- Les avantages obtenus par les approches BU et TD pour le processus de demande non-stationnaire sont plus élevés que ceux qui sont associés avec les processus stationnaires en termes de précision des prévisions.

## 7. Organisation de la thèse

Au terme de la position esquissée dans ce résumé, la thèse est structurée comme suit:

Dans le deuxième chapitre, nous présentons un état de l'art sur la prévision de la demande par l'agrégation. Différents types d'agrégation, c'est à dire l'agrégation temporelle et transversale sont discutées. L'effet de l'agrégation sur la structure du processus est décrit et enfin, des travaux menés sur la prévision de la demande en appliquant l'agrégation sont discutés.

Dans le troisième chapitre, l'effet de l'agrégation temporelle non-cumulée sur la prévision de la demande est examinée lorsque la série chronologique suit un processus stationnaire. Pour chacun des processus considéré dans cette étude, le MSE théorique est développé à la fois au niveau de la comparaison désagrégée et agrégée. Ensuite, les résultats

de MSE sont comparés afin d'identifier les conditions de supériorité de chaque approche. Puis, l'analyse de simulation est menée afin d'examiner les résultats théoriques à la suite de l'investigation empirique.

Nous présentons dans le quatrième chapitre, l'effet de l'agrégation transversale sur la prévision de la demande. On suppose que la série suit soit un processus stationnaire soit un non-stationnaire. L'évaluation analytique est d'abord considérée suivie d'une étude de simulation pour tester et valider les résultats théoriques. De plus, certaines hypothèses sont relaxées par rapport à l'analyse théorique. Les résultats sont complétés par une analyse empirique utilisant un ensemble de données réelles pour valider les résultats.

Enfin, dans la dernière partie de cette recherche, nous résumons les résultats de chaque chapitre et nous présentons les conclusions de cette thèse. Les implications managériales et les limites de la recherche sont décrites, avec les perspectives pour le travail réalisé dans le cadre de la thèse.

## Lexique

### A

AutoRegressive  
Integrated Moving  
Average (ARIMA)  
processes

Autoregressive  
moyenne mobile  
intégré

AutoRegressive  
Moving Average  
(ARMA) processes  
AutoRegressive (AR)  
processes  
Moving Average (MA)  
processes

Autoregressive  
moyenne mobile

Aggregation level  
Aggregate demand  
Aggregate level  
Aggregation approach  
ARCH

Autoregressive  
Moyenne mobile  
Niveau d'agrégation  
Demande agrégée  
Niveau agrégé  
Approche d'agrégation  
Conditionnelle  
hétéroscédasticité  
autorégressive

### B

Buckets of time

La tranche de temp

Bottom-up approach

approche ascendante

### C

Cross-sectional  
aggregation  
Customer Service,  
Cut-off point

Agrégation  
transversale  
Service client  
Point de rupture

Cross correlation  
Cumulative  $m$  step  
ahead forecast

Corrélation croisée  
Prévision cumulée de  
 $m$  période d'avenir

### D

Disaggregate demand

Demande désagrégée

Disaggregate  
level(subaggregate  
level)

Niveau désagrégé

Disaggregation  
Disaggregation weights

Désagrégation  
Le poids de désagrégation

Non-Stationary

Non-stationnaire

### F

Fast moving items

Articles à rotation  
rapide

Forecasting

Prévision

Forecast horizon

Horieon de prévision

Forecast accuracy

Indicateur de précision

### G

Stock control  
GARCH

Gestion de stock  
Conditionnelle  
hétéroscédasticité  
autorégressive  
généralisée

### H

High frequency time  
series

Les séries  
chronologique à haute  
fréquence

### I

INARIMA

Autoregressive  
moyenne mobile Entier

### L

Low frequency time  
series

Les séries  
chronologique à basse  
fréquence

### M

Mean Square Error  
(MSE)  
Master Production  
Scheduling  
Material/capacity  
planning

Erreur quadratique  
moyenne  
Programme directeur  
de production  
Planification du  
matériel / capacité

### N

Non-overlapping

Non-Cumulé

Non-aggregation approach	Approche de nonagrégation
<b>O</b>	
Overlapping	Cumulé
One step ahead forecast	Prévision d'une période d'avenir
<b>P</b>	
Procurement	Achats
Product family	Famille de produit
Pattern	Schéma
Practitioner	Professionnelle
Process parameters	Paramètres de processus
<b>S</b>	
Stationary	Stationnaire
Single Exponential Smoothing	Lissage exponentielle simple
smoothing constant	La constante de lissage
Stock Keeping Units	Unité de gestion de stock
Slow moving items	Articles à rotation lente
Spare part	Pièce détachée
Sales and Operations Planning	Planification des ventes et des opérations
Sale	Vente
<b>T</b>	
Temporal aggregation	Agrégation temporelle
Time series	Séries chronologique
Top-down approach	Approche descendante
<b>U</b>	
Uncertainty	Incertitude
<b>V</b>	
Variability	Variabilité

---