A note on the forecast performance of temporal aggregation

Rostami-Tabar, B. , Babai, M.Z. , Syntetos, A. and Ducq, Y.

CURVE is the Institutional Repository for Coventry University

http://curve.coventry.ac.uk/open

A Note on the Forecast Performance of Temporal Aggregation

Abstract: Earlier research on the effects of non-overlapping temporal aggregation on demand forecasting showed the benefits associated with such an approach under a stationary AR(1) or MA(1) processes for decision making conducted at the disaggregate level. The first objective of this note is to extend those important results by considering a more general underlying demand process. The second objective is to assess the conditions under which aggregation may be a preferable approach for improving decision making at the aggregate level as well. We confirm the validity of previous results under more general conditions and we show the increased benefit resulting from forecasting by temporal aggregation at lower frequency time units.

1. INTRODUCTION

Rostami-Tabar et al. [11] considered the effect of non-overlapping temporal aggregation on demand forecasting. They assumed that the disaggregate series follow either an Auto-Regressive process of order one, AR(1) or a first order Moving Average process, MA(1) and the procedure employed for extrapolation purposes is the Single Exponential Smoothing, SES. (For a survey on the application of exponential smoothing methods see Gardner [4]). They compared the variance of the forecast error (or equivalently, the mean square error (MSE)) obtained based on forecasting using the aggregate demand to that resulting from the consideration of the disaggregate data. Comparisons were performed at the original (disaggregate) demand level. In this case, the aggregation approach works as follows: first aggregate demand data in non-overlapping time buckets; then extrapolate requirements using SES at the aggregate demand level; and, finally, disaggregate the aggregate forecasts at the original frequency level to produce a one-step-ahead forecast. The disaggregation approach relies upon a straight one-step-ahead extrapolation using SES at the original frequency level (i.e. forecasting in the classical way).

The researchers concluded that performance improvements related to the aggregation approach are a function of the aggregation level, the smoothing constant, and the process parameters. They found that for high levels of positive auto-correlation in the original series, the aggregation approach may be outperformed by the classical one for both processes considered. In contrast, high levels of aggregation and low values of the SES smoothing constant lead to a superior performance of the aggregation approach.

The first objective of this note is to extend the work of Rostami-Tabar et al. [11] to a more general underlying demand process. We do so by assuming that the disaggregate series follow an Auto-Regressive Moving Average process of order one, ARMA(1,1). An ARMA model often fits demand time series significantly better than pure AR or MA models [3]. In addition, ARMA processes have been found to provides good fit for demands of long lifecycle goods such as fuel, food products, machine tools, etc [2, 9]. The results presented in this Note are more general than those presented by Rostami-Tabar et al. [11], as both the AR(1) and MA(1) processes are special cases of the ARMA(1,1) process.

The second objective is to derive results when performance is measured at the aggregate rather than disaggregate (original series) level. This is important in many operational management decisions such as inventory control, for example, where temporal aggregation considerations over the lead time (or lead time plus review period) drive replenishments. Aggregation over the prevalent aggregation level is a necessity and not an option [8, 10, 12, 16]. The aggregation level needs to match the forecast horizon and performance needs to be evaluated at that level. A key question then to be answered is: should we forecast at the original disaggregate level and then obtain aggregate forecasts or should we temporally aggregate demand and extrapolate directly at that level? That is, when comparisons are undertaken at the aggregate level, the aggregation approach works as follows: first aggregate demand data in non-overlapping time buckets; then extrapolate requirements using SES at the aggregate demand level. The disaggregation approach in contrast relies upon a straight one-step-ahead extrapolation using SES at the

original frequency level followed by the multiplication of that forecast (by the length of the aggregation level) to obtain a forecast at the aggregate level. So if at the original disaggregate level we have monthly data, and the lead time is, for example, 3 months, then the one-step-ahead monthly forecast needs to be multiplied by 3 (assuming stationary demand) in order to produce an aggregate lead time demand forecast.

Temporal aggregation is an intuitively appealing approach to reduce demand uncertainty and it has been shown, under certain conditions, to lead to performance improvements when forecasting is required at the original (disaggregate) one-step-ahead level. Extending these findings to multiple-steps-ahead or aggregation level estimates and under a general stationary framework assumption, and developing insights into the conditions under which aggregation may or may not add any value to the forecasting process, should be of great value to both the theory and practice of forecasting.

The remainder of this Note is organized as follows. In Section 2 we present the assumptions behind this work and we derive MSE expressions at both the disaggregate and aggregate level. In Section 3, the impact of the process and control parameters on the superiority of each approach is analyzed. The conditions that determine the comparative performance of the two approaches at both levels of comparison are determined in Section 4 followed by an empirical analysis, conducted in Section 5, and the conclusions of this work, offered in Section 6.

2. ASSUMPTIONS AND MSE DERIVATIONS

2.1. Notations and assumptions

The following notation is used for the remainder of the paper.

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

ε_t: Independent random variables for non-aggregated demand in period t, normally distributed with zero mean and variance σ^2

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

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

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

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

μ: Expected value of non-aggregated demand in any time period

We assume that the disaggregate demand series d_t follows an ARMA(1,1) process that can be mathematically written in period t by (1).

$$d_t = \mu(1 - \phi) + \varepsilon_t + \phi\, d_{t-1} - \theta\varepsilon_{t-1}. \tag{1}$$

When demand follows an ARMA(1,1) process the auto-covariance is [1]:

$$\gamma_k = Cov(d_t, d_{i,t-k}) = \begin{cases} \dfrac{(1 - 2\phi\theta + \theta^2)}{1 - \phi^2}\sigma^2 & k = 0 \\ \dfrac{(\phi - \theta)(1 - \phi\theta)}{1 - \phi^2}\sigma^2 & |k| = 1 \\ \phi^{k-1}\gamma_{i,1} & |k| > 1 \end{cases}, \tag{2}$$

where γ_k is defined as the auto-covariance of lag k.

For different combinations of the process parameters, the resulting underlying structure changes considerably. **Table 1** presents the auto-correlation structure for different process parameters; this helps to demonstrate how the process behaves and can be useful when interpreting the results of the forthcoming analysis.

Table 1: Auto-correlation of ARMA (1,1) process

Case	Process parameter	Auto-correlation
1	$0<\phi<1$, $-1<\theta<0$	Always positive , $0<$Auto-correlation lag1<1,
2	$-1<\phi<0$, for any θ	Oscillation between positive and negative values
3	$0<\phi<1$, $0<\theta<1$ and $\phi>\theta$	Always positive, $0<$Auto-correlation lag1<1
4	$0<\phi<1$, $0<\theta<1$ and $\phi<\theta$	Always negative, $-0.5<$Auto-correlation lag1<0

1

2.2. MSE derivation at the disaggregate level

In this section, the *MSE* of the forecasts resulting from the disaggregate and the aggregate demand data is derived when the comparison is undertaken at the disaggregate level. In other words, one step-ahead forecasts, f_t , are considered.

The MSE before aggregation, MSE_{BA}, can be written as follows:

$$MSE_{BA} = Var(d_t - f_t) = Var(d_t) + Var(f_t) - 2Cov(d_t, f_t), \tag{3}$$

As shown in Appendix A, the expression of MSE_{BA} is:

$$MSE_{BA} = \frac{\sigma^2}{1-0.5\alpha}\left(\frac{\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2} - \frac{\alpha(\phi-\theta)(1-\phi\theta)}{\left(1-\phi^2\right)\left(1-\phi+\alpha\phi\right)}\right). \tag{4}$$

The MSE after aggregation, MSE_{AA}, is derived in Appendix A (where F_T is the forecast produced using the aggregated data and F_T/m the forecast at the disaggregate level).

$$MSE_{AA} = Var\left(d_t - \frac{F_T}{m}\right) = Var(d_t) + \frac{1}{m^2}Var(F_T) - \frac{2}{m}Cov(d_t, F_T) = \gamma_0 + \frac{1}{m^2}Var(F_T) - \frac{2}{m}Cov(d_t, F_T). \tag{5}$$

This may be expressed as follows:

$$MSE_{AA} = \sigma^2 \times \left(\frac{1}{m^2}\left(\frac{\beta}{(2-\beta)}\left(\begin{array}{c} m\left(\frac{\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2}\right)+ \\ \left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right) \end{array} \right) + \frac{2\beta(1-\beta)}{(2-\beta)^2}\left(\frac{\left(\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\left(\sum_{k=1}^{m}k\phi^{k-1}+\sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right)}{\left(1-\phi^m+\beta\phi^m\right)} \right) \right) + \left(\frac{\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2}\right) - \frac{2}{m}\left(\frac{\beta\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)}{1-\phi^m+\beta\phi^m} \times \frac{1-\phi^m}{1-\phi} \right) \right). \tag{6}$$

2.3. MSE derivation at the aggregate level

In this section, the *MSE* for both the aggregation and the non-aggregation approaches is derived at the aggregate level. The MSE_{BA} for the comparison at the aggregate level is obtained as follows: Firstly, one step ahead demand forecasts are generated based on the *SES* method. Then, the results are multiplied by the aggregation level *m*. This results in a forecast (i.e. cumulative *m*-step-ahead estimate) at the aggregate level:

$$MSE_{BA} = Var(D_T - mf_t) = Var(D_T) + m^2 Var(f_t) - 2m Cov(D_T, f_t), \tag{7}$$

The expression of MSE_{BA} is derived in Appendix B as follows:

$$MSE_{BA} = \sigma^2 \left(\begin{array}{c} \left(\frac{m\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2}\right)+\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right) \\ +\frac{m^2\alpha}{2-\alpha}\left(\frac{1-2\phi\theta+\theta^2}{1-\phi^2}\right)+\frac{2m^2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)}\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \\ -\frac{2m\alpha\left[(1-\phi)+\phi\left(1-\phi^{m-1}\right)\right]}{(1-\phi)(1-\phi+\alpha\phi)}\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \end{array} \right) \tag{8}$$

Now, the *MSE* resulting from the aggregate data is considered. Disaggregate demand is first aggregated to get low frequency demand. Then, the aggregate forecasts are generated based on the *SES* forecasting method. The MSE_{AA} is defined as:

$$MSE_{AA} = Var(D_T - F_T) = Var(D_T) + Var(F_T) - 2Cov(D_T, F_T), \tag{9}$$

The MSE_{AA} is derived in Appendix B and it is as follows:

$$MSE_{AA} = \sigma^2 \left(\frac{2\left(\frac{m(1-2\phi\theta+\theta^2)}{1-\phi^2} + \left(\sum_{k=1}^{m-1} 2(m-k)\phi^{k-1} \right)\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \right)}{2-\beta} - \frac{2\beta\left(\sum_{k=1}^{m} k\phi^{k-1} + \sum_{k=2}^{m} (k-1)\phi^{2m-k} \right)}{(2-\beta)(1-\phi^m+\beta\phi^m)} \left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2} \right) \right). \tag{10}$$

3. IMPACT OF THE PARAMETERS ON THE PERFROMANCE

In this section the effect of the parameters on the ratio MSE_{BA}/MSE_{AA} is analyzed, which we use as a measure of the superiority of each approach.

Figure 1 shows the impact of the parameters on the ratio of MSE_{BA}/MSE_{AA} when the comparison is undertaken at the disaggregate level. This addresses the first objective of the research (as defined in page 1). In **Figure 1,** it is revealed that for positive values of θ and negative values of ϕ, the aggregation approach always yields more accurate forecasts than the non-aggregation one. However, when θ takes negative values and ϕ takes positive values, the comparative results are reversed. Additionally, when both θ and ϕ are positive and $\theta < \phi$, the non-aggregation approach also performs better than the aggregation one.

By referring to **Table 1**, it can be seen that the latter cases correspond to a high positive auto-correlation, not only for lag 1 but also for higher time lags. On the contrary, in the former case, the auto-correlation is not always positive; either it is negative or it oscillates between positive and negative values. Therefore, the outperformance of the non-aggregation approach can be attributed to the high positive auto-correlation values. As it can be seen in **Figure 1**, for highly positive values of auto-correlation, no level of aggregation may improve the accuracy of forecasts. This is generally true regardless of the control parameter values.

The results of this study generally confirm the previous work conducted by Rostami-Tabar et al. [**11**]. The results are very similar to the case of AR(1) demand process. However, when the aggregation approach works, there is a slightly increased benefit for an ARMA(1,1) process compared to the AR(1). The difference between the aggregation performance on the MSE under the ARMA(1,1) and the AR(1) process is less than 1%.

Figure 2 presents the impact of the parameters on the ratio of MSE_{BA}/MSE_{AA} when the comparison is undertaken at the aggregate level, an important scenario that has not been considered by Rostami-Tabar et al.[**11**]. This addresses the second objective of the research. The results in **Figure 2** show that when the aggregation level is high (higher values of *m*) the aggregation approach always outperforms the non-aggregation one. However, when the aggregation level is short (lower values of *m*), the superiority under concern is a function of the parameter values. In these cases, the non-aggregation approach performs better than the aggregation one when the auto-correlation is highly positive. Otherwise, the aggregation approach provides more accurate results.

The analysis also shows that for a fixed value of the smoothing constants, increasing the aggregation level improves the accuracy of the aggregation approach. However, the percentage improvement is very low (less than 1%). Additionally, for a fixed aggregation level *m* and smoothing constant before aggregation α, the performance of the aggregation approach decreases as the β value increases. This is valid at both levels of comparison (aggregate and disaggregate).
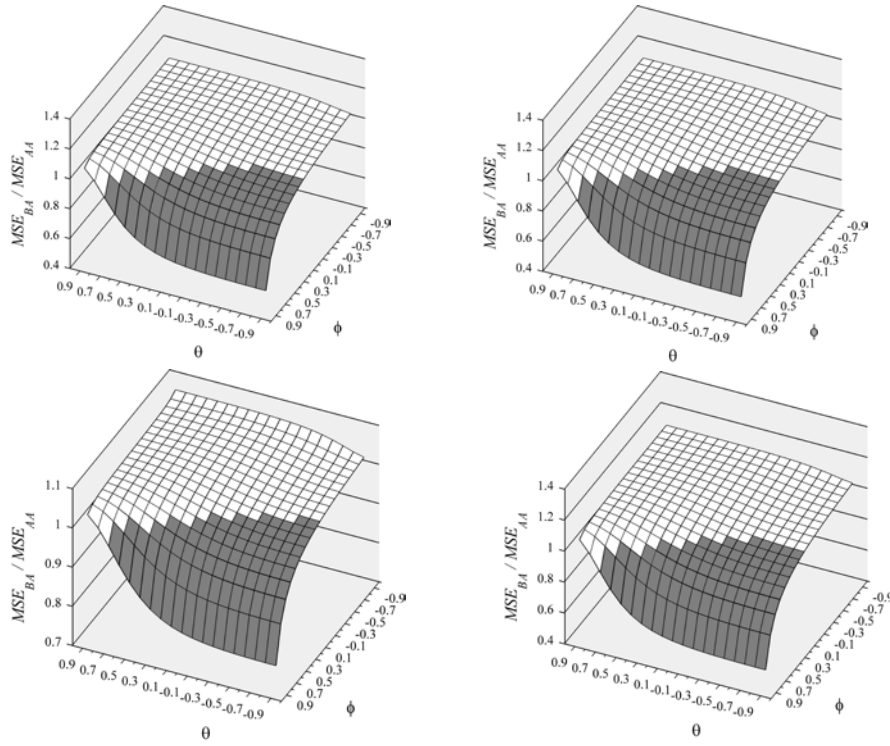
m=2   *m*=12

Figure 1: Impact of m, , θ, ϕ, α and β on the ratio of MSE at disaggregate level: $\alpha = 0.1, \beta = 0.01$ (top)
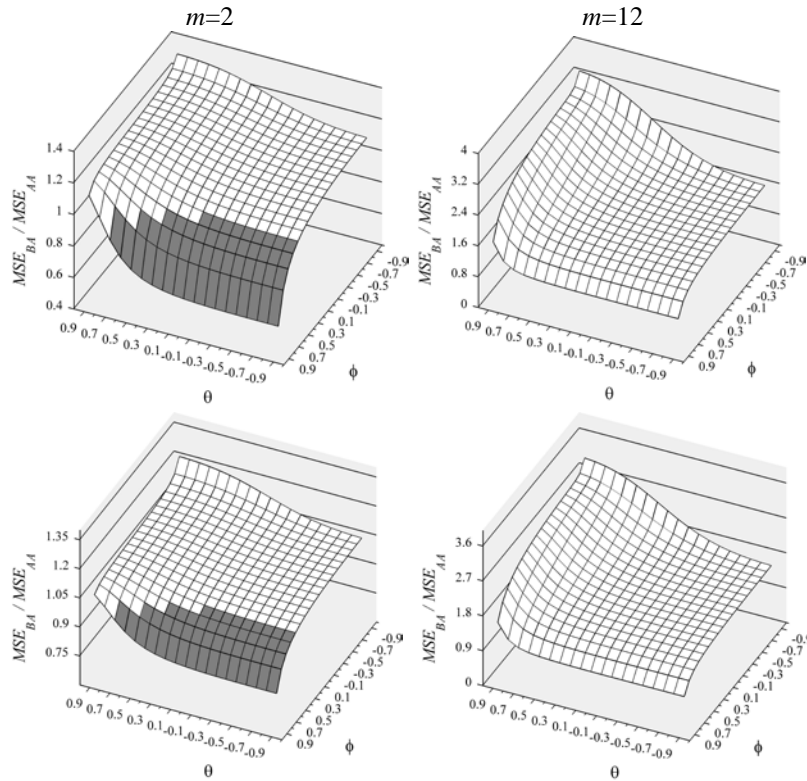
$\alpha = 0.1, \beta = 0.1$ (Bottom)

Figure 2: Impact of m, θ, ϕ, α and β on the ratio of MSE at aggregate level: $\alpha = 0.1, \beta = 0.01$ (top)

$\alpha = 0.1, \beta = 0.1$ (bottom)

What may be concluded at the end of this section is: i) the validity of the earlier results by Rostami-Tabar et al. [**11**] when the objective is to compare the forecasts at the disaggregate level; ii) the slightly improved performance of the aggregation approach for an ARMA(1,1) demand process compared to an AR(1) or MA(1) process. It is found that the aggregation approach always outperforms the non-aggregation one when the comparison is undertaken at the aggregate level and the aggregation level is high. However, for a lower aggregation level, the superiority is a function of the autocorrelation values.

The investigation reveals that the benefits of using the aggregation approach to produce aggregate (horizon) forecasts is more pronounced than its utilization at the disaggregate level. The further into the future an estimate is required, the forecast errors associated with the original disaggregate data become larger compared to the temporally aggregate series. In the next section, we determine theoretically the conditions under which each approach outperforms the other.

4. COMPARATIVE PERFORMANCE

Having conducted a sensitivity analysis, we now identify analytically the conditions under which each approach outperforms the other at both levels of comparison. To show the conditions under which the aggregation approach outperforms the non-aggregation approach, we set $MSE_{BA}/MSE_{AA} > 1$.

If the time series of the disaggregate demand follows an ARMA(1,1) process where $0 < \phi < 1$, then the conditions under which one approach outperforms the other at the disaggregate level of comparison can be obtained. These conditions are summarized in the selection procedure discussed in Appendix C.

THEOREM 1: If the time series of the disaggregate demand follow an ARMA(1,1) process where $-1 < \phi \leq 0$, and the comparison is undertaken at the disaggregate level, then the following conditions determine the superiority of each approach :

- If $\beta < \beta_1$, the aggregation approach provides more accurate forecasts.

- If $\beta = \beta_1$, both approaches perform equally.

- Otherwise, the non-aggregation approach works better.

where β_1 is as defined in (C-4).
PROOF: The proof of Theorem 1 is given in Appendix D.

Next, the conditions under which each approach provides more accurate forecasts at the aggregate level of comparison are determined. The ratio of MSE_{BA}/MSE_{AA} for comparison at the aggregate level is determined by dividing (8) into (10). If the time series of the disaggregate demand follow an ARMA(1,1) process where $0 < \phi < 1$, and the comparison is undertaken at the aggregate level, then the superiority conditions of each approach can be obtained. These conditions are summarized in a selection procedure presented in Appendix E.

THEOREM 2: If the time series of the disaggregate demand follows an ARMA(1,1) process where $-1 < \phi \leq 0$, and the comparison is conducted at the aggregate level, then:

- If $\beta < \beta_1$, the aggregation provides more accurate forecasts.

- If $\beta = \beta_1$, both approaches perform equally.

- Otherwise, the non-aggregation approach works better.

where β_1 is defined in (E-3).
PROOF: The proof of Theorem 2 is given in Appendix F.

Theorems 1 and 2 show that when the autoregressive and the moving average parameters satisfy $-1 < \phi \leq 0$, then for a given value of the smoothing constant, α, and the aggregation level, m, there is always a value of β for which the aggregation approach provides more accurate forecasts.

5. EMPIRICAL ANALYSIS

In this section, we assess the empirical validity of the main theoretical findings of this research. The demand dataset available for the purposes of this research consists of weekly sales data over a period of two years for 1,798 stock keeping units (SKUs) of a major European supermarket located in Germany; 5.1% of the SKUs (91 series) were identified as ARMA(1,1) by using the *R* package.

In **Table 2,** we summarize the characteristics of the SKUs relevant to our study by indicating the estimated parameters for the ARMA(1,1) process.

Table 2: Processes present in the empirical data set, ARIMA(1,0,1) process

θ intervals	ϕ intervals	Average of θ	Average of ϕ	Average lag1Auto-correlation	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

We must remark that θ and ϕ do not cover the entire theoretically feasible range and the auto-correlation of the data under consideration is positive.

Figure 3a presents the results of the empirical analysis when the comparison is undertaken at the disaggregate level. **Figure 3a** indicates that for all values of the aggregation level *m,* the MSE_{BA} is lower than the MSE_{AA} when the optimal smoothing constant values α and β are used. Therefore, for all values of the aggregation level the non-aggregation approach outperforms the aggregation one. It should be noted that in order to facilitate presentation purposes the results are expressed based on the *RMSE* (root mean square error) rather than the *MSE*. The forecast error reduction can be as high as 8% and this is in agreement with our theoretical findings; the real data (please see **Table 2)** is associated with positive auto-correlation not only for lag 1 but for longer time lags as well.
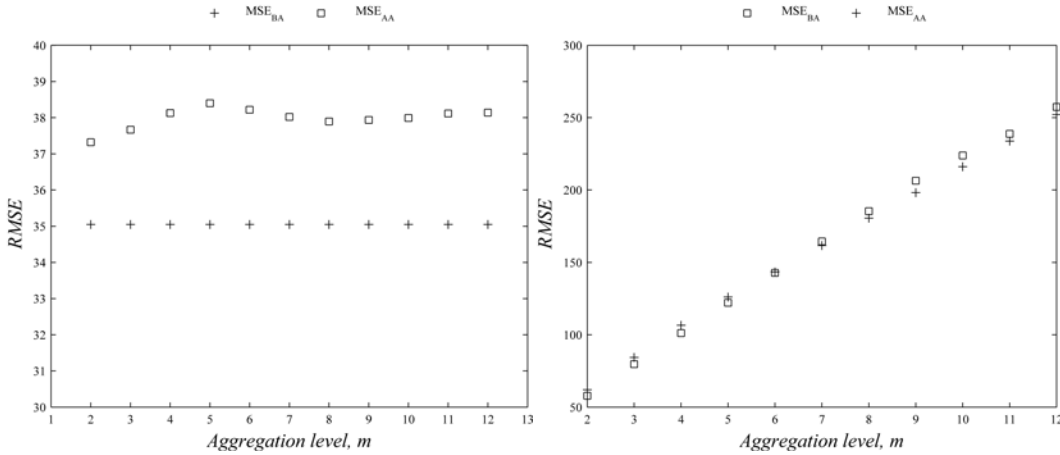

a) Comparison at the disaggregate level   b) Comparison at the aggregate level

Figure 3: Empirical results

Figure 3b shows the results of the empirical analysis when the comparison is undertaken at the aggregate level. The results indicate that for an aggregation level $m \leq 6$, the MSE_{BA} is smaller than the MSE_{AA}, i.e. the non-aggregation approach performs better. However, as the aggregation level increases, $m > 6$, the aggregation approach should be preferred. Therefore, the empirical results reveal that when the disaggregate demand follows an ARMA(1,1) process with positive auto-correlation, there is a cut-off point of the aggregation level below which the non-aggregation approach performs better and above which the comparative performance is reversed. Hence, the empirical analysis confirms the results of the theoretical evaluation at both levels of comparison. However, it should be noted that even when the

aggregation approach outperforms the non-aggregation one, the difference is quite small and overall it can be concluded from the empirical investigation that there is no clear advantage of the aggregation approach.

6. IMPLICATIONS AND CONCLUDING REMARKS

We derived MSE expressions to facilitate the identification of conditions under which non-overlapping temporal aggregation may add value in the forecasting process under the presence of stationary ARMA(1,1) demand. Performance was evaluated at both the disaggregate and aggregate level assuming extrapolation based on a Single Exponential Smoothing (SES) procedure. The main findings of this note can be summarized as follows:

- In the first part of the study, the previous work of Rostami-Tabar et al. [11] is entirely confirmed. In fact, when the disaggregate series follows an ARMA(1,1) process and the comparison is conducted at the disaggregate level, then for high values of positive auto-correlation in the original series the aggregation approach is outperformed by the non-aggregation one. For these values, no level of aggregation improves the performance of the aggregation approach. However, when the autocorrelation value is negative, less positive or oscillates between positive and negative values, the aggregation approach is preferred. These results are very similar to those observed in the case of an AR(1) process but there is more benefit resulting from the aggregation of an ARMA(1,1) demand process compared to an AR(1) process.
- The aggregation approach, when associated with higher aggregation levels and lower smoothing constant values after aggregation, β, provides more accurate forecasts. This is true for comparisons at both the disaggregate and aggregate level.
- We find that there is more benefit associated with the aggregation approach when producing aggregate than disaggregate forecasts.
- We reveal that when the comparison is undertaken at the aggregate level (forecast horizon), then the superiority of each approach is a function of the autocorrelation and the aggregation level. If the aggregation level is low, then: i) for highly positive auto-correlation values the non-aggregation approach may perform better; ii) for less positive or negative auto-correlation values, the aggregation approach is preferred. However, if the aggregation level is high, then the aggregation approach may outperform the non-aggregation one regardless of the process and control parameters. This is a very important result for practitioners. As the horizon over which forecasts are required increases, the forecast errors associated with the original data become larger and the utilization of temporally aggregated data becomes indispensable.
- Due to the high positive autocorrelation of the dataset used in the empirical investigation, overall there is no clear advantage of the aggregation approach.

As far as the next steps of research are concerned, and in addition to the suggestions provided by Rostami-Tabar et al. [11], further work into the following areas would appear to be merited: i) the interface between temporal and cross-sectional aggregation [6, 7]; ii) the impact of temporal aggregation on forecasting non-stationary processes, in particular trended series [5, 13]; iii) the extension of the analysis discussed here to other real world datasets to increase confidence in the empirical validity of our findings.

APPENDIX A: MSE DERIVATION-COMPARISON AT DISAGGREGATE LEVEL

In order to calculate the MSE_{BA} at the disaggregate level, we first need to calculate the covariance between the disaggregate demand and its forecast, which is given by:

$$Cov(d_t, f_t) = \alpha\gamma_1 + \alpha(1-\alpha)\phi\gamma_1 + \alpha(1-\alpha)^2\phi^2\gamma_1 + \ldots = \frac{\alpha\gamma_1}{1-\phi+\alpha\phi}. \tag{A-1}$$

The variance of the disaggregate forecast is calculated as follows:

$$Var(f_t) = \frac{\alpha\gamma_0}{2-\alpha} + \frac{2\alpha(1-\alpha)\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)} \ , \tag{A2}$$

Substituting (2), (A-1) and (A-2) in (3) in conjunction with the fact that $\gamma_0 = Var(d_{t-k})$ results in the MSE_{BA} being as follows:

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

When the disaggregate series follows an ARIMA(1,1) process, the aggregate series also follows an ARMA(1,1) process but with different parameter values [**15**] as follows:

$$\gamma_0' = m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right) \tag{A-4}$$

$$\gamma_1' = \gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right) \tag{A-5}$$

$$\phi' = \phi^m \tag{A-6}$$

Now to obtain the MSE_{AA} at disaggregate level, we need to calculate i) the variance of aggregate forecast ii) the covariance between the disaggregate demand and the aggregate forecast. The latter can be calculated as follows:

$$Cov(d_t, F_T) = \beta Cov(d_t, D_{T-1}) + \beta(1-\beta)Cov(d_t, D_{T-2}) + \beta(1-\beta)^2 Cov(d_t, D_{T-2}) + ... \tag{A-7}$$

By substituting the aggregate demand, D_{T-k} into (A-7) we have:

$$Cov(d_t, F_T) = \beta(\gamma_1 + \gamma_2 + ... + \gamma_m) + \beta(1-\beta)(\gamma_{m+1} + \gamma_{m+2} + ... + \gamma_{2m}) + \beta(1-\beta)^2(\gamma_{2m+1} + \gamma_{2m+2} + ... + \gamma_{3m}) + \cdots \tag{A-8}$$

By substituting (2) into (A-8) and making some simplifications, we have

$$Cov(d_t, F_T) = (1 + \phi + ... + \phi^{m-1})\times(\beta\gamma_1 + \beta(1-\beta)\phi^m\gamma_1 + \beta(1-\beta)^2\phi^{2m}\gamma_1 + ...) \tag{A-9}$$

By doing some simple calculation we get

$$Cov(d_t, F_T) = \frac{\beta\gamma_1}{1-\phi^m + \beta\phi^m}\times\frac{1-\phi^m}{1-\phi} \tag{A-10}$$

The covariance between the aggregate demand and its forecast and the variance of the aggregate forecast can be derived directly by using (A-1) and (A-2) by replacing the disaggregate parameters by the aggregate ones as the process still remains ARMA(1,1). By doing so we get:

$$Cov(D_T, F_T) = \frac{\beta\gamma_1'}{1-\phi' + \beta\phi'} \tag{A-11}$$

$$Var(F_T) = \frac{\beta\gamma_0'}{2-\beta} + \frac{2\beta(1-\beta)\gamma_1'}{(2-\beta)(1-\phi' + \beta\phi')} \tag{A-12}$$

Now by substituting (A-10) and (A-12) into (5) and then (A-4), (A-5) (A-6), and (2) into that, the MSE of the forecast after aggregation is given as follows:

$$MSE_{AA} = \frac{1}{m^2}\times\frac{\beta\left(\begin{array}{l}m\left(\frac{(1-2\phi\theta+\theta^2)}{1-\phi^2}\sigma^2\right)+\\\left(\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\sigma^2\right)\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right)\end{array}\right) + \frac{2\beta(1-\beta)\left(\left(\frac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\sigma^2\right)\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right)}{(2-\beta)(1-\phi^m + \beta\phi^m)}}{2-\beta} + \left(\frac{(1-2\phi\theta+\theta^2)}{1-\phi^2}\sigma^2\right) - \frac{2}{m}\left(\frac{\beta\gamma_1}{1-\phi^m + \beta\phi^m}\times\frac{1-\phi^m}{1-\phi}\right). \tag{A-13}$$

APPENDIX B: MSE DERIVATION-COMPARISON AT AGGREGATE LEVEL

In order to calculate the MSE_{BA} at the aggregate level, first we need to calculate the covariance between the aggregate demand and the disaggregate forecast, which is given by:

9

$$Cov(D_T, f_t) = \frac{\alpha\gamma_1\left[(1-\phi)+\phi\left(1-\phi^{m-1}\right)\right]}{(1-\phi)(1-\phi+\alpha\phi)} \tag{B-1}$$

1   By considering that $\gamma_0' = Var(D_T)$ and substituting (A-2), (B-1) and then (A-4), (A-5), (A-6), and
2   finally (2) into (7), the MSE_{BA} is as follows:

$$MSE_{BA} = \sigma^2 \left(\begin{array}{l} \left(\dfrac{m\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2}\right) + \left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\left(\displaystyle\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right) \\[3mm] + \dfrac{m^2\alpha}{2-\alpha}\left(\dfrac{1-2\phi\theta+\theta^2}{1-\phi^2}\right) + \dfrac{2m^2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)}\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \\[3mm] - \dfrac{2m\alpha\left[(1-\phi)+\phi\left(1-\phi^{m-1}\right)\right]}{(1-\phi)(1-\phi+\alpha\phi)}\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \end{array} \right) \tag{B-2}$$

3   Now by considering that $\gamma_0' = Var(D_T)$ and substituting (A-11) and (A-12) in (9), the MSE_{AA} at
4   aggregate level is obtained as follows:

$$MSE_{AA} = \frac{2\gamma_0'}{2-\beta} - \frac{2\beta\gamma_1'}{(2-\beta)(1-\phi'+\beta\phi')}. \tag{B-3}$$

5   By substituting (A-4), (A-5) and (A-6) into (B-3), the following equation is given:

$$MSE_{AA} = \frac{2\left(m\gamma_0 + \gamma_1\left(\displaystyle\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right)}{2-\beta} - \frac{2\beta\left(\gamma_1\left(\displaystyle\sum_{k=1}^{m}k\phi^{k-1} + \displaystyle\sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right)}{(2-\beta)(1-\phi^m+\beta\phi^m)}. \tag{B-4}$$

6   Finally, by substituting (2) into (B-4), the MSE_{AA} becomes:

$$MSE_{AA} = \frac{2\left(\dfrac{m\left(1-2\phi\theta+\theta^2\right)}{1-\phi^2} + \left(\displaystyle\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\right)}{2-\beta} - \frac{2\beta\left(\displaystyle\sum_{k=1}^{m}k\phi^{k-1} + \displaystyle\sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)}{(2-\beta)(1-\phi^m+\beta\phi^m)}\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right). \tag{B-5}$$

7   **APPENDIX C: SELECTION PROCEDURE-COMPARISON AT DISAGGREGATE LEVEL**

8   By the ratio of $MSE_{BA}/MSE_{AA} > 1$, the quadratic function given by (C-1) should be negative

$$\begin{aligned} &\left(\begin{array}{l} \phi^m(1-\phi)\left(m\gamma_0 + \gamma_1\left(\displaystyle\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\[3mm] -2(1-\phi)\left(\gamma_1\left(\displaystyle\sum_{k=1}^{m}k\phi^{k-1} + \displaystyle\sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) + 2m\left(1-\phi^m\right)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{array} \right)\beta^2 \\[4mm] &+\left(\begin{array}{l} \left(1-\phi^m\right)(1-\phi)\left(m\gamma_0 + \gamma_1\left(\displaystyle\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\[3mm] +2(1-\phi)\left(\gamma_1\left(\displaystyle\sum_{k=1}^{m}k\phi^{k-1} + \displaystyle\sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) \\[3mm] -4m\left(1-\phi^m\right)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)\left(1-\phi^m\right)\eta \end{array} \right)\beta - 2m^2(1-\phi)\left(1-\phi^m\right)\eta \end{aligned} \tag{C-1}$$

9   where

$$\eta = \frac{\alpha\gamma_0}{2-\alpha} - \frac{2\alpha\gamma_1}{(2-\alpha)(1-\phi+\alpha\phi)} \tag{C-2}$$

10   Moreover, by investigating the sign of (C-1) we can obtain the conditions under which
11   MSE_{BA}/MSE_{AA} is smaller than, equal to, and greater than one. Now, we check whether the quadratic
12   function (C-1) has real roots. To do so, we define the discriminant Δ of (C-1) as follows

$$\Delta = \left[\begin{array}{l} \left(1-\phi^m\right)\left(1-\phi\right)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\ + 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) \\ - 4m\left(1-\phi^m\right)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{array} \right]^2 + 8\left(m^2(1-\phi)(1-\phi^m)\eta\right) \left[\begin{array}{l} \phi^m(1-\phi)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\ - 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) \\ + 2m\left(1-\phi^m\right)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{array} \right], \quad \text{(C-3)}$$

1   Now by using the fact that $-1 < \theta < 1$, $0 < \phi < 1$, $0 < \alpha < 1$ and $m \geq 2$, the values of Δ can be
2   obtained. If $\Delta < 0$ it means (C-1) has no real roots and if $\Delta > 0$ it means (C-1) has two real roots called
3   β_1 and β_2, where

$$\beta_1 = \cfrac{-\left[\begin{array}{l} \left(1-\phi^m\right)\left(1-\phi\right)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\ + 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) - 4m\left(1-\phi^m\right)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{array} \right] + \sqrt{\Delta}}{2\left[\begin{array}{l} \phi^m(1-\phi)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) - 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) + \\ 2m\left(1-\phi^m\right)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{array} \right]}, \quad \text{(C-4)}$$

$$\beta_2 = \cfrac{-\left[\begin{array}{l} \left(1-\phi^m\right)\left(1-\phi\right)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\ + 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) - 4m\left(1-\phi^m\right)\gamma_1 - 2m^2\phi^m(1-\phi)\eta + m^2(1-\phi)(1-\phi^m)\eta \end{array} \right] - \sqrt{\Delta}}{2\left[\begin{array}{l} \phi^m(1-\phi)\left(m\gamma_0 + \gamma_1\left(\sum_{k=1}^{m-1}2(m-k)\phi^{k-1}\right)\right) \\ - 2(1-\phi)\left(\gamma_1\left(\sum_{k=1}^{m}k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\right) + 2m\left(1-\phi^m\right)\gamma_1 + m^2\phi^m(1-\phi)\eta \end{array} \right]}. \quad \text{(C-5)}$$

4   It is known that the sign of the (C-1) between the two roots β_1 and β_2 is opposite to the sign of A,
5   where A is the sign of the coefficient of β^2, otherwise it is that the same as the sign of A.
6   If the discriminant $\Delta < 0$, there are no real roots for (C-1), therefore the sign of (C-1) is equivalent to
7   the sign of A. We can show that when $\Delta < 0$, A is always negative, consequently (C-1) is negative which
8   means that MSE_{BA}/MSE_{AA} is smaller than one.
9   However, If $\Delta > 0$, (C-1) has two different roots β_1 and β_2. By investigating the sign of β_1, β_2 and
10   A, we can determine the sign of (C-1) and consequently the performance superiority of each approach.
11   The superiority conditions of each approach can be obtained by following the selection procedure:
12   1. The procedure begins by calculating Δ defined in (C-3). If $\Delta < 0$ then the non-aggregation approach
13   is always superior, otherwise the values of β_1 and β_2 defined in (C-4) and (C-5) are calculated.
14   2. If $\beta_2 \in (0,1)$, the value of β_1 is calculated and according to the values of β_1 and β_2 the following
15   results are obtained:
16   ▪ If $\beta_2 < \beta < \beta_1$, then the aggregation approach works better.
17   ▪ If $\beta = \beta_1 = \beta_2$, then both approaches are identical.
18   ▪ If $\beta > \beta_1$ or $\beta < \beta_2$, then disaggregate strategy works better. Otherwise, go to 3.
19   3. If $\beta_2 \notin (0,1)$, we calculate the value of β_1:
20   ▪ If $\beta < \beta_1$, then the aggregation approach works better.
21   ▪ If $\beta = \beta_1$, then both approaches are identical.
22   ▪ If $\beta > \beta_1$, then non-aggregation approach works better.

APPENDIX D: PROOF OF THEOREM 1

2 In this case, the process parameters satisfies $-1 < \theta < 1$ and $-1 < \phi \le 0$, It can be shown that for these

3 parameter values Δ defined at (C-3) is always positive and $\beta_2 < 0$ or $\beta_2 > 1$.

4 Now by considering β_1, β_2 and A that is positive for $\theta < 0$ and negative for $\theta > 0$, the sign of (C-1) is

5 determined. So we have

6 ▪ If $\beta_2 < 0$ and $\beta_1 > 0$, then (C-1) is negative in the interval $[\beta_2, \beta_1]$ and it is positive outside this

7 interval.

8 ▪ If $\beta_2 > 1$, we can show that $0 < \beta_1 < \beta_2$ and (C-1) is positive in the interval $[\beta_1, \beta_2]$ and it is

9 negative outside this interval

10 From the above expressions, the following results can be obtained:

11 ▪ If $\beta < \beta_1$, then $MSE_{BA}/MSE_{AA} > 1$.

12 ▪ If $\beta = \beta_1$, then $MSE_{BA}/MSE_{AA} = 1$.

13 ▪ Otherwise, $MSE_{BA}/MSE_{AA} < 1$.

14 **APPENDIX E: SELECTION PROCEDURE - COMPARISON AT AGGREGATE LEVEL**

15 Considering $MSE_{BA}/MSE_{AA} > 1$ is equivalent to having the quadratic function (E-1) negative, which

16 subsequently is equivalent to

$$\left(-\phi^m \Psi\right)\beta^2 + \left(2\phi^m \Psi - \Psi(1-\phi^m) - 2\phi^m \eta + 2\xi\right)\beta + 2(1-\phi^m)(\Psi - \eta) \tag{E-1}$$

17 where

$$\Psi = \begin{vmatrix} \left(\dfrac{m(1-2\phi\theta+\theta^2)}{1-\phi^2}\right) + \left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\left(\displaystyle\sum_{k=1}^{m-1} 2(m-k)\phi^{k-1}\right) \\ + \dfrac{m^2\alpha}{2-\alpha}\left(\dfrac{1-2\phi\theta+\theta^2}{1-\phi^2}\right) + \dfrac{2m^2\alpha(1-\alpha)}{(2-\alpha)(1-\phi+\alpha\phi)}\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \\ - \dfrac{2m\alpha\left[(1-\phi)+\phi(1-\phi^{m-1})\right]}{(1-\phi)(1-\phi+\alpha\phi)}\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right) \end{vmatrix}$$

18

$$\eta = \left(\dfrac{m(1-2\phi\theta+\theta^2)}{1-\phi^2} + \left(\sum_{k=1}^{m-1} 2(m-k)\phi^{k-1}\right)\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)\right)$$

19

$$\xi = \left(\sum_{k=1}^{m} k\phi^{k-1} + \sum_{k=2}^{m}(k-1)\phi^{2m-k}\right)\left(\dfrac{(\phi-\theta)(1-\phi\theta)}{1-\phi^2}\right)$$

20

21 For the quadratic function given by (E-1), the value of the discriminant Δ and the roots β_1 and β_2

22 can be defined as follows:

$$\Delta = \left(2\phi^m \Psi - \Psi(1-\phi^m) - 2\eta\phi^m + 2\xi\right)^2 + 8\phi^m \Psi(1-\phi^m)(\Psi - \eta) \tag{E-2}$$

$$\beta_1 = \dfrac{\left(2\phi^m \Psi - \Psi(1-\phi^m) - 2\eta\phi^m + 2\xi\right) + \sqrt{\Delta}}{2\phi^m \Psi} \tag{E-3}$$

$$\beta_2 = \dfrac{-\left(2\phi^m \Psi - \Psi(1-\phi^m) - 2\eta\phi^m + 2\xi\right) + \sqrt{\Delta}}{-2\phi^m \Psi} \tag{E-4}$$

23 By following the same procedure as discussed in Appendix C, the superiority conditions of each

24 approach can be obtained by following the selection procedure:

25 1. The procedure begins by calculating Δ defined in (E-2), If $\Delta < 0$ then the non-aggregation approach

26 is always superior, otherwise the values of β_1 and β_2 defined in (E-3) and (E-4) are calculated.

27 2. If $\beta_2 \in (0,1)$, the value of β_1 is calculated and according to the values of $\beta1$ and $\beta2$ the following

28 results are obtained:

29 ▪ If $\beta_2 < \beta < \beta_1$, then the aggregation approach works better.

30 ▪ If $\beta = \beta_1 = \beta_2$ then both approaches are identical.

- If $\beta > \beta_1$ or $\beta < \beta_2$ then disaggregate strategy works better. Otherwise, go to 3.
3. If $\beta_2 \notin (0,1)$, we calculate the value of β_1:
  - 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.

APPENDIX F: PROOF OF THEOREM 2

In this case we have $-1 < \theta < 1$ and $-1 < \phi \le 0$. It can be shown that for these parameter values Δ defined at (E-2) is always positive and β_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 C 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.

REFERENCES

[1]. G. Box, G. M. Jenkins and G. Reinsel, *Time series analysis: Forecasting & control*, John Wiley and sons, 2008.

[2]. S. Chopra and P. Meindl, *Supply chain management*, Prentice-Hall, New Jersey, 2001.

[3]. T. T. H. Duc, H. T. Luong and Y.-D. Kim, *A measure of bullwhip effect in supply chains with a mixed autoregressive-moving average demand process*, European Journal of Operational Research **187** (2008), no. 1, 243-256.

[4]. E. S. Gardner, Jr., *Exponential smoothing: The state of the art—part ii*, International Journal of Forecasting **22** (2006), no. 4, 637-666.

[5]. E. S. Gardner, Jr. and E. McKenzie, *Forecasting trends in time series*, Management Science **31** (1985), no. 10, 1237-1237.

[6]. M. Kalchschmidt, R. Verganti and G. Zotteri, *Forecasting demand from heterogeneous customers*, International Journal of Operations & Production Management **26** (2006), no. 6, 619-638.

[7]. N. Kourentzes, F. Petropoulos and J. R. Trapero, *Improving forecasting by estimating time series structural components across multiple frequencies*, International Journal of Forecasting, **30** (2014), no. 2, 291-302.

[8]. M. Mohammadipour and J. E. Boylan, *Forecast horizon aggregation in integer autoregressive moving average (inarma) models*, OMEGA **40** (2012), no. 6, 703-712.

[9]. S. Nahmias, *Production and operations analysis*, Irwin Publishers, Boston, MA., 1993.

[10]. K. Nikolopoulos, A. A. Syntetos, J. Boylan, F. Petropoulos and V. Assimakopoulos, *An aggregate-disaggregate intermittent demand approach (adida) to forecasting : An empirical proposition and analysis*, Journal of the Operational Research Society **62** (2011), no. 3, 544-554.

[11]. B. Rostami-Tabar, M. Z. Babai, A. Syntetos and Y. Ducq, *Demand forecasting by temporal aggregation*, Naval Research Logistics (NRL) **60** (2013), no. 6, 479-498.

[12]. J. W. Taylor, *Density forecasting of intraday call center arrivals using models based on exponential smoothing*, Management Science **58** (2012), no. 3, 534-549.

[13]. ---, *Exponential smoothing with a damped multiplicative trend*, International Journal of Forecasting **19** (2003), no. 4, 715-725.

[14]. G. C. Tiao, *Asymptotic behaviour of temporal aggregates of time series*, Biometrika **59** (1972), no. 3, 525-531.

[15]. W. W. S. Wei, *Time series analysis: Univariate and multivariate methods*, Addison-Wesley Publishing Company, Boston, USA, 2006.

[16]. G. Zotteri and M. Kalchschmidt, *A model for selecting the appropriate level of aggregation in forecasting processes*, International Journal of Production Economics **108** (2007), no. 1-2, 74-83.