EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1550 / MAY 2013

FORECASTING FISCAL TIME SERIES USING MIXED FREQUENCY DATA

Stylianos Asimakopoulos, Joan Paredes
and Thomas Warmedinger

In 2013 all ECB publications feature a motif taken from the €5 banknote.

Acknowledgements
We thank seminar participants at the European Central Bank for helpful comments and suggestions. We are also grateful to Christian Schumacher and Eric Ghysels for providing support, comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank.

Stylianos Asimakopoulos
University of Glasgow, Adam Smith Business School; e-mail: s.asimakopoulos.1@research.gla.ac.uk

Joan Paredes
European Central Bank; e-mail: joan.paredes@ecb.europa.eu

Thomas Warmedinger
European Central Bank; e-mail: thomas.warmedinger@ecb.europa.eu

Address	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
Postal address	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Internet	http://www.ecb.europa.eu
Fax	+49 69 1344 6000

ISSN	1725-2806 (online)
EU Catalogue No	QB-AR-13-047-EN-N (online)

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=2264101.

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

Abstract

Given the increased importance of fiscal monitoring, this study amends the existing literature in the field of intra-annual fiscal data in two main dimensions. First, we use quarterly fiscal data to forecast a very disaggregated set of fiscal series at annual frequency. This makes the analysis useful in the typical forecasting environment of large institutions, which employ a "bottom-up" or disaggregated framework. Aside from this practical type of consideration, we find that forecasts for total revenues and expenditures via their subcomponents can actually result more accurate than a direct forecast of the aggregate. Second, we employ a Mixed Data Sampling (MiDaS) approach to analyze mixed frequency fiscal data, which is a methodological novelty. It is shown that MiDaS is the best approach for the analysis of mixed frequency fiscal data compared to two alternative approaches. The results regarding the information content of quarterly fiscal data confirm previous work that such data should be taken into account as it becomes available throughout the year for improving the end-year forecast. For instance, once data for the third quarter is incorporated, the annual forecast becomes very accurate (very close to actual data). We also benchmark against the European Commission's forecast and find the results fare favorably, particularly when considering that they stem from a simple univariate framework.

Keywords: Fiscal policy, Mixed frequency data, Short-term forecasting, Aggregated vs. disaggregated forecast.

JEL-Classification: C22, C53, E62, H68.

1 Non-technical summary

The current economic turmoil has led to an increase in uncertainty when forecasting annual budgetary executions and their subcomponents. For fiscal monitoring it is essential to assess the quality of intra-annual data and their implications for the end-year result. Several papers have highlighted the importance of intra-annual fiscal data for forecasting annual budgetary outturns either through directly forecasting deficits or indirectly, via the two main subcomponents (total revenues and expenditures) (e.g. Onorante *et al.*, 2010, Paredes *et al.*, 2009, Pedregal and Pérez, 2010 and Pérez, 2007). Although these results are important and can signal risks to the budgetary execution, they are not directly useful when preparing disaggregated fiscal forecasts. To this end, it is important to assess also the effect of intra-annual fiscal data on the annual outturn of disaggregated series. Against that background, we extend the existing literature on the implications of using intra-annual fiscal data for forecasting not only deficits, total revenues and expenditures, but also their subcomponents.

Moreover, we employ a Mixed Data Sampling (MiDaS) approach to analyze mixed frequency fiscal data, which as far as we are aware has never been used before to forecast individual euro area countries fiscal time series. We compare this approach with two alternatives and show that MiDaS performs best for the analysis of mixed frequency fiscal data.

Our results confirm that quarterly fiscal data include significant information and they should be taken into account as they become available throughout the year. It is shown that the simple (univariate) model constructed in this paper for forecasting individual fiscal series can produce very accurate forecasts of the annual outcome. The results support the use of quarterly fiscal data and demonstrate the usefulness of the MiDaS approach when dealing with mixed frequency fiscal data.

We also contribute to the literature regarding the question of aggregate versus disaggregate forecasts. To this end we use the forecast obtained from subcomponents to indirectly forecast total revenues and total expenditures. We find that on the expenditure side there is a consistent forecast improvement when the disaggregated approach is used. This may relate to the fact that expenditures are more difficult to forecast as they depend on discretionary government decisions. Therefore, timely and good quality data are a powerful tool to improve these variable forecasts. However, on the revenue side results are not as consistent, although there are for many countries also significant benefits from the disaggregated approach.

2 Introduction

The sovereign debt crisis has increased the importance of fiscal monitoring and forecasting. It is therefore crucial to assess the relevance of incoming intra-annual data, distinguishing in particular news from noise, and their implication for the annual budgetary outturn. The usefulness of intra-annual fiscal data has been shown in many studies (e.g. Onorante *et al.*, 2010, Paredes *et al.*, 2009, Pedregal and Pérez, 2010 and Pérez, 2007). The focus of the existing literature is mainly on the extraction of news from higher frequency data for the forecast of the current-year budgetary outturn in terms of the overall deficit. This is useful as such and can be used to signal risks to budgetary executions. Hughes Hallett *et al.* (2012) show how forecasts from higher frequency data can improve fiscal surveillance and planning.

For the purpose of preparing fiscal forecasts it is important to assess disaggregated intra-annual data because fiscal forecasts by large institutions are normally prepared in a "bottom-up" approach. The aim of this paper is thus to assess the news content of quarterly fiscal data releases and the implications for the annual outturn of those series. Besides this, there are also implications for the quarterly profile, in particular for the remaining quarters of a given year. This is less interesting from a purely fiscal perspective, but it is highly relevant if such quarterly fiscal profiles are used as input for short-term macroeconomic projections. The primary purpose of this analysis is thus to extent the existing literature towards the implications of intra-annual fiscal data also for a more disaggregated set of variables.

A second purpose of the analysis concerns the methodology. We employ the Mixed Data Sampling approach (MiDaS) as proposed by Ghysels *et al.* (2004) for the time series analysis at hand, in which regressand and regressor are sampled at different frequencies. This mixed data sampling approach has so far mainly been used for volatility predictions for financial sector data (e.g. Forsberg and Ghysels (2006) and Ghysels *et al.* (2006)). MiDaS has also been used in forecasting macroeconomic variables using intra-annual data. For example, Bai *et al.* (2009), Clements and Galvao (2008, 2009) and Kuzin *et al.* (2011) use monthly data to improve quarterly forecast of macroeconomic time series. Following the same procedure, Andreou *et al.* (2010) and Ghysels and Wright (2009) use daily financial data to nowcast macroeconomic data of monthly or quarterly frequency.

The advantage of this approach compared to alternative ones, such as State Space and mixed frequency VAR models, which make use of the Kalman filter, is that MiDaS is more parsimonious and less sensitive to specification errors due to the use of non-linear lag polynomials.

To the best of our knowledge, there is no other application of the MiDaS approach for fiscal time series analysis except for a recent working paper by Ghysels and Ozkan (2012). Ghysels and Ozkan (2012) use MiDaS together with intra-annual macroeconomic US data and indicators to forecast deficits, total revenues and total expenditures. We deviate from that work in the sense that we forecast directly fiscal aggregates and their subcomponents, and also the aggregates indirectly through their subcomponents. It is therefore the first time that a forecast is performed on individual fiscal time series using mixed frequency data. Although our work comprises a much bigger country sample, it suffers from a smaller time-span sample regarding the series under consideration.

In addition, we contribute to the literature that favors forecasting indirectly aggregates via their subcomponents. In particular, after a forecast has been performed for the subcomponents of revenues and expenditures, we aggregate and compare their forecast with the total revenues and expenditures direct forecast. Lütkepohl (2010) in his survey indicates that a disaggregated forecast of the components and then aggregating the results can lead to better forecasts in terms of lower mean squared forecast errors. This is due to the richer information contained in the components.

The aggregated versus disaggregated approach has also been assessed in the context of GDP forecasting. In particular, Baffigi *et al.* (2004) use euro area countries as a whole and find that with bridge models the aggregated forecast of the total GDP is more accurate than aggregating the forecast of its components. However, when individual countries forecast is assessed, Marcellino *et al.* (2003) show that it is better to forecast the euro area GDP via aggregating the forecast of individual countries (disaggregated approach). Moreover, Perevalov and Maier (2010) indicate that forecasting the US economic activity through the expenditure components may report slightly improved forecast.

The paper is organized as follows. Section 3 provides a description of the econometric models under consideration. Section 4 describes the data. Section 5 describes the methodology of the empirical application. Section 6 has the results of the case study. Section 7 contains an analysis of aggregated versus disaggregated forecasting of aggregates. Section 8 concludes the paper.

3 Model specification

This section provides an overview of the various models that will be used in the empirical analysis later on. The first model is the simple aggregation approach in which the high-frequency variables are transformed to low frequency by simply taking their average. The second model is the unrestricted mixed frequency data analysis. Using this model requires no assumption regarding the high frequency variables. However, other issues may arise, like the parameter proliferation issue. The benchmark model is the mixed frequency data sampling model (MiDaS) that makes use of a distributed lag polynomial, which is data driven and non-linear in order to transform the high-frequency data into low frequency.

3.1 Flat weight aggregation approach

The most simple case of dealing with mixed frequency data is to aggregate the high frequency data and then take their average. This approach implies equal weights on each quarter. However, if the true weighting scheme is not that of equal weights, the average estimation will lead to biased estimators.

In more detail, assuming that Y_{t+1}^A is the annual time series and that X_t^Q is the quarterly time series, the distributed lag regression applied is the following:

$$Y_{t+1}^A = \beta_0 + \beta_1 X_t^A + u_{t+1} \tag{1}$$

where $X_t^A = \sum_{i=1}^{N_Q} \frac{1}{N_Q} X_{i,t}^Q = \left(X_{N_Q,t}^Q + X_{N_Q-1,t}^Q + X_{N_Q-2,t}^Q + X_{N_Q-3,t}^Q \right) / N_Q$, is the annual time series obtained from the quarterly data, with N_Q denoting the number of quarters within a year.

In addition, assuming that ω_i are the weights assigned to each quarter (i), and using the quarterly lag operator L_Q^i we can re-write the above equation as:

$$Y_{t+1}^A = \beta_0 + \bar{\beta}_1 \sum_{i=1}^{M} \omega_i L_Q^i X_{i,t}^Q + u_{t+1}^A$$

Comparing this equation with equation (1), taking the aggregation scheme of the quarters into account, the following expression can be obtained:

$$Y_{t+1}^A = \beta_0 + \bar{\beta}_1 X_t^A + \bar{\beta}_1 \sum_{i=1}^{N_Q} \left(\omega_i - \frac{1}{N_Q} \right) L_Q^i X_{i,t}^Q + u_{t+1}^A$$

Therefore, if the true weighting scheme is not the "equal / average" weighting scheme, the simple OLS regression will have biased estimators because of the omitted regressor (the third term in the above equation). As a consequence, the slope coefficient will be biased because of the misspecified model and this affects the forecasting accuracy of the model.

3.2 Unrestricted regression

As an alternative model the high frequency variable can be directly related to the low frequency variable without the need of aggregation (see e.g. Foroni *et al.*, 2011):

$$Y_{t+1}^{A} = \beta_0 + \sum_{i=0}^{N_Q-1} \beta_j X_{N_Q-i,t}^{Q} + u_{t+1} \tag{2}$$

The above equation is estimated using ordinary least squares (OLS). The advantage of this approach is that it does not make any assumption on the weights that should be attached to each quarter (unrestricted) and that it can directly estimate mixed frequency data. This estimation is called Unrestricted Mixed frequency Data Sampling (U-MiDaS).

The number of parameters that need to be estimated in this approach increases significantly in comparison to the previous case. In particular, there are five parameters to be estimated when dealing with annual/quarterly data.[1] However, this number can further increase if monthly data are being used instead of quarterly data, or if several lags of each quarter are incorporated. Therefore, the aforementioned approach (U-MiDaS) suffers from the parameter proliferation issue.

3.3 Mixed frequency Data Sampling approach (MiDaS)

Ghysels *et al.* (2006) proposed the Mixed frequency Data Sampling (MiDaS) approach where the parameter proliferation issue can be avoided, and no assumption is required regarding the attached weights to high frequency variables. There are only three parameters to be estimated in the single variable distributed lag model case, which is invariant with respect to the frequency or the lag length of explanatory variable. This is because the MiDaS regression is based on distributed lag polynomials to ensure a parsimonious specification. But the lag polynomials are not linear and therefore MiDaS regression is estimated using non-linear least squares (NLS).

Denoting the high frequency data (quarterly data in this case) with X_t^Q and the low frequency data (annual data) with Y_t^A, the typical MiDaS regression is the following:

$$Y_{t+1}^{A} = \mu + \beta \sum_{j=0}^{q_X^Q-1} W\left(L^{N_Q};\theta\right) X_{t-j}^{Q} + \varepsilon_{t+1}$$

[1] One coefficient for each quarter and one for the constant.

where, $W\left(L^{N_Q};\theta\right)$ defines the weights attached to each lag of the quarterly data. Also, L^{N_Q} is a simple quarterly lag operator and θ is a composition of two parameters that determine the curvature of the weighting scheme.

Specifically, the distributed lag polynomial is given as:

$$W\left(L^{N_Q};\theta\right)X_t^Q=\sum_{j=0}^{N_Q-1}\omega_j\left(\theta_X^Q\right)X_{t-j}^Q$$

This is the expression of the lag polynomial that will determine the effect (weight) of the explanatory variable on the dependent variable. Ghysels *et al.* (2007) propose various weighting schemes. They showed that the exponential Almon lag polynomial is the most general weighting scheme as it is very flexible and can take many shapes. This polynomial needs only two parameters, $\theta=(\theta_1,\theta_2)$, to be calibrated using the data. The weights are therefore purely data driven and no prior assumption is required.

The expression of the exponential Almon lag polynomial, $\omega_j\left(\theta_X^Q\right)$, is the following:

$$\omega_j\left(\theta_1,\theta_2\right)=\frac{\exp\{\theta_1 j+\theta_2 j^2\}}{\sum_{j=1}^{m}\exp\{\theta_1 j+\theta_2 j^2\}}$$

Therefore, the following distributed lag (DL) model can be obtained using the MiDaS method and the exponential Almon lag polynomial.

$$Y_{t+1}^A=\beta_0+\beta_1\sum_{j=0}^{q_X^Q-1}\sum_{i=0}^{N_Q-1}\omega_{i+j*N_Q}\left(\theta^Q\right)X_{N_Q-i,t-j}^Q+\varepsilon_{t+1} \tag{3}$$

This method is called DL-MiDaS(q_X^Q).[2] Note that q_X^Q denotes the number of lags of the high frequency variable after it has been transformed to low frequency using the lag polynomial.

It is possible to augment the above equation using lagged variables of the dependent variable. In this case the following autoregressive version of the DL-MiDaS (ADL-MiDaS(p_Y^A,q_X^Q)) can be obtained:[3]

$$Y_{t+1}^A=\beta_0+\sum_{i=0}^{p_Y^A-1}\rho_i Y_{t-i}^A+\beta_1\sum_{j=0}^{q_X^Q-1}\sum_{i=0}^{N_Q-1}\omega_{i+j*N_Q}\left(\theta^Q\right)X_{N_Q-i,t-j}^Q+\varepsilon_{t+1} \tag{4}$$

However, the addition of autoregressive regressors might cause discontinuities in the impulse response function of X_t^Q on Y_{t+1}^A(see Ghysels *et al.*,

[2]Note that throughout the paper we also refer to DL-MiDaS simply as MiDaS.

[3]Where p_Y^A denotes the number of lags of the low frequency variable (annually in this case) that included in the regression.

2007). To address this issue Clements and Galvao (2008) propose a common factor restriction for adding the autoregressive component and ensure in this way smooth impulse responses. However, in this paper the autoregressive component will not be applied. This is because the high frequency variables are the disaggregated version of the low frequency variables in the univariate MiDaS model applied in this analysis, and adding an autoregressive component would therefore cause multicollinearity issues.[4]

Another important characteristic of the MiDaS approach is that the slope coefficient β, can be easily obtained from the regression as the weights attached to the high frequency data are normalized and sum to one. In addition, MiDaS is much more flexible than a flat-weighting scheme since it can nest the equal weighting scheme by setting $\theta_1 = \theta_2 = 0$. MiDaS can also take seasonality into account by attaching the appropriate weight to each lagged regressor.[5]

3.4 Nowcasting with MiDaS

Using high frequency data for forecasting low frequency data can lead to a more accurate forecast in the case where data from within the forecast period are utilized. Especially during periods of economic turmoil, like the current period, intra-annual data releases can improve forecast accuracy.

The MiDaS regression can take this new information into account and perform the nowcasting. In this case the DL-MiDaS presented earlier in equation (3) becomes:

$$Y_{t+1}^{A} = \beta_0 + \beta_1 \sum_{j=0}^{q_X^Q - 1} \sum_{i=0}^{N_Q - 1} \omega_{i+j*N_Q}\left(\theta^Q\right) X_{N_Q-i,t+1-j-s/4}^{Q} + \varepsilon_{t+1} \qquad (5)$$

where s is the forecast horizon in quarters. Note that the time period for the high frequency variable, X_t^Q, is not t any more but $t+1$ and depends on the quarterly information released within the year of forecast, as it is determined from s.

[4] However, an ADL-MiDaS approach is implemented and compared with DL-MiDaS to verify that the inclusion of an autoreggresive term does not improve the accuracy of the forecast. The results reported in the appendix show that the ADL-MiDaS leads to higher RMSFEs than the DL-MiDaS for almost every case study. Therefore, not including an autoreggresive term does not affect the results reported in this work.

[5] For verifying that seasonality is not an issue in MiDaS regression the data have been de-seasonalized and then re-estimated with MiDaS. The forecast reported from this method is very similar (very similar RMSFE) with the forecast reported when vintage data are used.

When $s < 4$ information of the current year is being used and nowcasting is being performed. For example, $s = 3$ denotes a forecast horizon equal to 3 quarters ahead. The model will in such case perform nowcasting since it is going to use data from the first quarter of the year to update next period's forecast. It is also possible to set $s > 4$. The model will in that case perform a forecast using all the available data from more than one year before the desired forecasting year. This way a forecast can be performed for multiple periods ahead.

4 Data

The data are formed by a very disaggregated set of annual and quarterly fiscal variables. We firstly transform the data to render them stationary. This is achieved by taking the differences of the natural logarithms.

Specifically, the fiscal data used for all the case studies are the vintages of the Government Finance Statistics database for quarterly data as published by Eurostat and the vintages of the DG-ECFIN AMECO database, which includes annual data and forecasts. In particular, the data include total revenues (TOR), direct taxes (DTX), indirect taxes (TIN), social contributions (SCT), total expenditures (TOE), social benefits other than in kind (THN), interest payments (INP), subsidies (SIN), compensation of employees (COE) and government investment (GIN).

Moreover, the period that is covered varies depending on the country and the availability of the data. For example, for Belgium and France fiscal data are available from 1991q1 to 2012q1 and for the rest of the countries the data sample is smaller, from 1999q1 to 2012q1.[6]

The main focus of this paper is the forecast of next period (i.e. period $t + 1$) fiscal variables (end-of-year forecast) taking into account the higher frequency data for the same period. In other words, quarterly fiscal data will be used as they become available in each quarter, to perform nowcasting of the annual fiscal series. Nowcasting in the literature refers mostly to forecast updates using high frequency data. For example, if the aim is to forecast the annual growth rate of total revenues (TOR) using quarterly data, the nowcasting approach updates the 2012 forecast for TOR growth rate using data of the explanatory variables of either the first and/or the second and/or the third quarter in 2012. The Mixed Data Sampling (MiDaS) approach

[6] The other countries are: Austria, Germany, Finland, Ireland, Italy, Luxembourg, Netherlands, Portugal, Slovenia and Spain.

incorporates available higher frequency data within the forecast period to nowcast.

There are some specific features that should be born in mind about the data. The available annual and quarterly data are revised every 6 months. This is related to the semi-annual (April-October) reporting obligation of annual Government Finance Statistics (GFS) by Member States to Eurostat, as stated within the context of the Excessive Deficit Procedure (EDP). The data are usually revised backwards up to three years. For example, when new data become available in 2011q3, the data are revised backwards approximately until 2008q3. When performing forecasts only information available at that point in time will be used. These data are called end-of-sample vintage data (EndVint) and include a combination of first announcements and revised data. This is the type of data included throughout the analysis.

However, Koening *et al.* (2003) suggest that real-time vintage data (RTVin) are better because they are not revised. However, fiscal data are only available in RTVin format from 2005q2. Therefore, it is not possible to construct the RTVin database using only data without any revision, while having at the same time a sufficiently large sample for the econometric analysis.

5 Methodology

In a first step, the three aforementioned models will be compared in terms of their forecasting performance. The benchmark model is the distributed lag MiDaS regression, as in equation (3). The benchmark model will be compared with a simple aggregation scheme, also called flat weighting scheme, as in equation (1) (named flat-weight) and with the Unrestricted-MiDaS, equation (2) (named U-MiDaS). The comparison of the different models is based on the Root Mean Squared Forecast Error (RMSFE). Several boxplots[7] are constructed to ease the comparison of all model forecasting performance.

A model forecast comparison will also be conducted when nowcasting.[8] This means that information available within the forecast period will be used. Overall, there will be four boxplots reported for each country. The first one includes the RMSFE without any use of information within the forecast period. Therefore, the first boxplot of each country makes no use of quarterly information within the forecasted year.[9]

[7]The boxplots can be found in the appendix.

[8]Nowcasting in the literature refers mostly to forecast updates using available high frequency data within the same forecast period.

[9]For that reason the first boxplot of each country is labeled as April (Q0).

In contrast, the other three boxplots will include the results of RMSFE from the nowcasting exercise. In this case the boxplot on the top right corner, labeled as July (Q1), takes into account quarterly information of the fiscal variable up until the first quarter (inclusive) of that current year. Following the same concept, the other two boxplots, labeled as Oct. (Q2) and Jan. (Q3), take into account quarterly information of fiscal data up until the second and third quarter respectively of the year that we are forecasting.

5.1 Lag-length determination

It is also important to determine the lag length of the explanatory variables because under-parameterizing the model might lead to a significant increase of the RMSFE (e.g. Götz *et al.*, 2012). In the case of DL- MiDaS the lag-length will be determined through the Bayesian Information Criteria (BIC) with a minimum of 6 and a maximum of 12 lags. This means that quarterly information of at least $1^{1/2}$ years and up to 3 years will be included in the regression. In any case, even if more lags than necessary would be included, the lag polynomial would assign a zero weight to the unimportant lags and the forecast accuracy would not be affected.

For the flat-weight model two lags are included in the regression, which means that two years of intra-annual information will be taken into account for the forecast.[10] However, the quarterly data lag length for the Unrestricted MiDaS cannot be larger than six due to the small sample.

5.2 Benchmark model vs European Commission forecast

Finally, the forecasting ability of the benchmark model (DL-MiDaS) will be compared with the forecast reported by the European Commission (EC). As it turns out, even though the DL-MiDaS model is purely univariate and therefore very simple, it can still produce very accurate forecasts. Timmermann (2006) points to the fact that it is not possible for an individual model to outperform all others at each point in time because such forecasting models are thought of as local approximations. Also, Stock and Watson (2004) suggest that a combination of forecasts using many different variables and models can result in a much more accurate and robust forecast than an individual model.

[10] The flat-weight model has also been examined with 3 years of information included in the regression (3-lags) and the forecast accuracy did not change significantly.

It is therefore inaccurate to state that any model (included the univariate model used in this paper) is the best model to forecast fiscal variables. However, it can be concluded that high-frequency fiscal data appear to have a significant bearing on forecasting performance. As the year advances and subsequently more quarterly data can be incorporated into the MiDaS model its forecast improves, and even sometimes outperforms the forecast from the EC. This can be seen from the boxplots. When nowcast is being performed, the boxplot for MiDaS is much closer to zero compared to the relevant boxplot of the European Commission. However, the most relevant comparison between European Commission and MiDaS is for the case where the second quarter is taken into account (the boxplots labeled as Oct. (Q2)). At that time the European Commission updates its forecast and takes into account the available information until that quarter, as in the MiDaS approach. For that particular quarter the results are mixed, with MiDaS reporting more accurate forecast for at least half of the countries for some variables. Nevertheless, the forecast improvement is very pronounced when the third quarter is taken into account.

Overall, the above results indicate that the MiDaS model can be very promising in terms of forecasting performance and could be incorporated as one of the possible forecasting models.

6 Empirical Results

6.1 Comparison of the forecasting ability

The following figures show the root mean squared forecast errors (RMSFEs) for the different models and the European Commission.[11] Boxplots closer to zero indicate better forecasts. Note that at this stage only the overall forecast from each model is compared for each quarter and not the individual time series. Therefore, the following boxplots show the evolution of the RMSFE of the various models and variables when new quarterly data becomes available as the forecasted year advances.[12]

[11] The boxplots can be found in the appendix.

[12] The boxplot with title "April (Q0)" includes only information as released in April. In April annual together with the last quarter of previous year fiscal data are released . As a result there is no benefit stemming from quarterly data at that point in time. The second boxplot with title "July (Q1)" means that in July there is the first quarter release of the current year and will be taken into account when updating the annual forecast. At the boxplots wih title "Oct. (Q2)" and "Jan. (Q3)" the second quarter and the third quarter are included, respectively. As a result in each boxplot the forecast is being

For Belgium and France, where we have a larger sample available, the estimation period starts in 1991q1. For the rest of the countries, it starts in 1999q1. The estimation period ends in 2007q4 for all countries. We use then data until 2011q4 for a rolling estimation of one year ahead forecasts for the period 2008q1-2011q4. This way there are four different forecasts/nowcasts calculated in each of the four years for each quarter.

The RMSFE is obtained from each forecast/nowcast giving 4 different RMSFEs for 10 fiscal variables for each country.[13]

The boxplot is then obtained by setting initially the 10 different RMSFEs in an ascending order. The median is represented by the red line inside the box. The black lines outside of the box represent the top and bottom 25% of the observations. Finally, inside the box there are the remaining 50% observations. In some cases there are also a few outliers which are noted with a red cross outside of the box. The closer the boxplot is to zero the better is the forecast from that approach.

The RMSFE results indicate that the MiDaS approach outperforms the flat-weighting and U-MiDaS approach in every case. The MiDaS approach is therefore among them the best model for using intra-annual fiscal data to forecast the annual outturn of those series .

Moreover, the European Commission's forecast has been included in the boxplots for comparison. It is worth noting that the information that the European Commission is using for their forecast is by far richer than the univariate MiDaS model applied here. It could normally not be expected that a univariate model (like the one constructed in this paper) could improve upon the European Commission's forecast.

However, from the boxplots can be seen that for at least half of the countries included in this case study, the MiDaS model has actually smaller RMSFE than European Commission's forecast when nowcasting is performed. In addition, since the RMSFEs reported in the boxplots apply to one period ahead forecasts for four different years, it hints that the combination of MiDaS and quarterly information can improve the fiscal data forecasting accuracy.

[Figures 1-12 here]

---

updated accordingly taking the new quarterly fiscal data release into account as soon as they become available.

[13] Note that in the appendix the Diebold-Mariano test results are provided for testing the statistical significance of the difference in RMSFEs between the various approaches.

6.2 How to improve our forecast in real-time? Keep on incorporating quarterly data

As mentioned previously, the boxplots indicate that high frequency fiscal data (quarterly in this case) contains important information that should be taken into account when fiscal variables are forecasted. It is shown that the forecast reported from the MiDaS approach improves in most of the cases when the nowcasting is performed. It can thus be concluded that when high frequency fiscal data become available within the forecast period they should be included to update the forecast (nowcasting).[14]

The table below shows how the information contained with a new release of quarterly fiscal data is assessed. In particular, it is examined whether the inclusion of one additional quarter will improve the forecast performance in terms of the RMSFE. For example, the inclusion of the first quarter will improve the forecast of TOR in 11 out of 12 countries. However, the inclusion of the second quarter will improve the forecast only in 5 out of 12 countries. As mentioned before, this might be the result of past data revisions, which occur always when Q2 (in October) and Q4 (in April) data are published.

Table.1 Univariate MiDaS improvement

	TOR	DTX	TIN	SCT		
Q_1	11	6	11	11		
Q_2	5	5	9	11		
Q_3	10	10	12	9		
	TOE	THN	INP	SIN	COE	GIN
Q_1	10	10	7	7	10	9
Q_2	9	9	9	9	11	6
Q_3	10	10	10	9	10	9

From the above table several important conclusions can be drawn. First of all, it is shown that the inclusion of a new quarter systematically improves the forecast accuracy of individual fiscal variables both on the revenue and expenditure side. This holds for most of the countries with the exception of Q2 for TOR and DTX, when data revisions occur.

When there are no data revisions the inclusion of a new quarter release improves the forecast significantly. This is obvious with the inclusion of the third quarter, where the forecast improves for at least nine out of twelve countries.

[14] Note that so far only the overall forecasting ability of MiDaS model has been assessed. At the next section an analysis of the individual fiscal time series is provided and the forecasting/nowcasting of each variable will be compared with the actual data.

These results indicate clearly that quarterly data contains significant news and improves thus the end-year forecast.

6.3 Benchmarking MiDaS forecast accuracy

Since it has been concluded that quarterly fiscal data contain significant information, it is also important to assess how well the forecast compares to the actual data. It has been shown in the literature (e.g. Artis and Marcellino, (2001) and Keereman (1999)) that the forecast reported by the European Commission(EC) is very accurate, so it serves as a natural benchmark for the MiDaS forecast.

The EC's fiscal forecasts take into account intra-annual information, macroeconomic variables and models as well as experts' beliefs. It is therefore a forecast composed from many different variables and indicators which can be expected to increase its accuracy. In contrast, the MiDaS model employs only historical information from the same variable under consideration, i.e. it is a very simple univariate model. Nevertheless, to qualify the accuracy of MiDaS forecasts they will be compared with the forecasts reported from the European Commission.

Table.2 Univ. MiDaS vs Eur. Comm.

	TOR	DTX	TIN	SCT
Q_0	0	1	0	2
Q_1	1	2	4	2
Q_2	0	0	3	4
Q_3	1	4	8	7

	TOE	THN	INP	SIN	COE	GIN
Q_0	1	0	5	4	2	1
Q_1	1	2	7	4	4	5
Q_2	1	1	7	5	4	3
Q_3	5	6	7	7	7	4

The above table shows the evolution of MiDaS forecasts for each variable and quarter. For instance, for indirect taxes (TIN) there is in Q_0 (April) (without using any intra-annual information) no country in which the RMSFE from MiDaS is lower than the EC's one. However, when the first quarterly data become available for Q_1 (July) there are four countries where the forecast is more accurate when using MiDaS univariate model for the specific variable. However, this result is against the DG-ECFIN AMECO April release, which does not incorporate information on this quarter.

If we compare forecast performances including Q2 data, which the DG-ECFIN AMECO release also incorporates, it is to be highlighted that on the expenditure side Univ. MiDaS seems to be able to extract important news that might not be fully exploited by the EC. For example, Univ. MiDaS is able to improve the EC forecast for 7 countries for interest payments (INP).

The importance of quarterly information is specially relevant on the expenditure side of fiscal forecasts as many times there are no clear macroeconomic variables to which expenditure items can be linked to. This is less the case on the revenue side, where macroeconomic tax bases are clearly defined. For instance, indirect taxes can be expected to follow private consumption dynamics.

Finally, the most significant improvement comes in the last quarter (Q_3), released in January of the following year, where for most of the variables and for half of the countries MiDaS model can improve the accuracy of the forecast by taking into account only the last quarter release of data. The comparison is this time again not fully fair as the DG-ECFIN Autumn release is not able to incorporate this info, which is only published in January of the following year.

6.4 Country-specific analysis and increasing the sample size

The country-specific graphs below confirm the conclusions of the above tables, i.e. that the inclusion of quarterly fiscal data releases improves the forecast accuracy.[15] This is shown from the negative slope of the MiDaS Univariate curve. Moreover, since the graphs are country specific, conclusions can also be drawn about the quality of the quarterly fiscal data in each country. For countries where this is not the case this could be a clear sign quality lack in their quarterly GFS data. This is one of the reasons mentioned by fiscal economists who have been adamant to use quarterly GFS to date. The improvement in the collection of these data would improve greatly the capability of nowcasting tools for country monitoring.[16] In any case, our analysis does not support in general the case against using quarterly data.

[15] The figures can be found in the appendix.

[16] Note that these graphs are obtained after adding the RMSFEs obtained earlier for each variable. For example, in April, where the annual fiscal data are released, the point in graph is obtained after adding the RMSFEs for each variable as they have been calculated in the previous subsection. In July the first quarter of fiscal data are released and being taken into account for the calculation of the RMSFEs and the update of the annual series forecast (exactly as it was the case in the boxplots). As in the boxplots, in October the second quarter of fiscal data is released, whereas in January the third quarter is released.

As it has been stated earlier, for France and Belgium the sample size is almost double in comparison with the rest of the countries. In addition, the graphs indicate that for these two countries the MiDaS forecast is very accurate (lower RMSFE than the EC), even after the first quarter data release. Taking these two facts into account, it could be concluded that the sample size is as expected important for improving the MiDaS forecast accuracy.

In order to further assess this point, an additional sample of observations that was available for some of the countries has been taken into account for three variables (DTX, TIN, THN), indicated in the graphs as "MiDaS_ext ". The additional sample goes back to 1991 which is exactly the starting point of all the variables for Belgium and France. As a result, when the "MiDaS_ext " line on the graph is below the MiDaS Univariate line it means that the inclusion of the extra sample has improved the overall accuracy of the forecast.

The graphs indicate that in every country for which the additional sample was available the forecast accuracy improves. This reinforces the argument on the small sample bias, and that the accuracy of the forecast that MiDaS reports becomes more accurate with a larger sample size. As a result, the MiDaS performance will improve as new data become available.[17]

[Figures 13-18 here]

7 Forecasting aggregates using subcomponents

In this section we examine whether forecasting total revenues and total expenditures through a disaggregated approach using their subcomponents will result in a more accurate forecast of the aggregate compared to forecasting the aggregate directly. For that purpose the relevant share of the components relative to the corresponding aggregate series and also the RMSFE, as constructed previously, has been taken into account.[18]

The difference with the boxplots is that here the RMSFEs for each variable are being added to calculate the aggregate RMSFE for each quarter.

[17] Also, in the appendix are provided the results from the Diebold-Mariano test for the countries where the additional sample is available to test whether the improvement from the additional sample can result in statistically significant more accurate forecast.

[18] Note that through this analysis we have not included the additional sample for the three variables as described in the previous section. This way the comparison between disaggregate and aggregate forecasting is more consistent since data for aggregate series are available only from 1999 (apart from Belgium and France).

The graphs below show the improvement of the forecast (lower RMSFE) when aggregates are forecasted using subcomponents.[19] In addition, the EC's forecast of the aggregates has been included so as to assess the magnitude of the improvement.[20]

In the graphs there appears to be a consistent improvement of total expenditures forecast through the indirect approach. For most of the countries this improvement is significant in magnitude and in some cases the RMSFE falls below the EC's RMSFE. On the revenue side, results are not as consistent but again the benefits from the disaggregate forecast are significant for most of the countries in terms of lower RMSFE.[21]

[Figures 19-30 here]

8 Conclusions

In this paper, we employ a Mixed frequency Data Sampling (MiDaS) approach to assess the importance of quarterly fiscal data when forecasting their annual outturn. We compare the MiDaS model with two alternatives and we find that MiDaS is among them the best approach for analyzing mixed frequency data. The results regarding the significance of the quarterly fiscal data indicate that they should be taken into account as they become available throughout the year so as to update the annual fiscal forecast(nowcasting). In particular, when the third quarter of fiscal data becomes available the updated forecast is very accurate (very close to actual data). For the very few countries where this is not the case, this could hint to a lack of quality in their quarterly GFS data. This is one of the reasons mentioned by fiscal economists who have been adamant to use quarterly GFS to date. The improvement in the collection of these data would improve greatly the capability of nowcasting tools for country monitoring.

Furthermore, the results indicate that the longer span of fiscal data are available (e.g. in the case of Belgium and France we have 8 more years of

[19] Note that on the vertical axis the root mean squared forecast error is measured.

In addition, on the horizontal axis Q0 indicates that at this point in time there is no information regarding quarterly data of the current year (equivalent to April in previous graphs). Q1 is the equivalent of July and shows that the first quarter of the current year has been released and will be taken into account to update the forecast. In addition, Q2 and Q3 are the equivalent of October and January, respectively in the previous graphs.

[20] The figures can be found in the appendix.

[21] In the appendix can be found the results from the Diebold-Mariano test regarding the statistical significance of the results.

quarterly fiscal data than we have for the rest of the countries) the more accurate MiDaS forecast appears to be. In order to assess these results we include a larger span of quarterly data that was available for three variables and for seven countries. This exercise shows that indeed the additional sample improves the accuracy of the forecast in terms of a lower RMSFE.

Finally, we also examine whether the forecast of total revenues and total expenditures via their subcomponents (disaggregated approach) can result in a more accurate forecast than directly forecasting those series (aggregated approach). We find that on the expenditure side there is a consistent improvement of the forecast when the disaggregated approach is used (in almost every country the RMSFE of total expenditures is lower). This could be related to the fact that the expenditure side of the public accounts is more difficult to forecast as it depends greatly on discretionary government decisions and therefore timely and good quality data are a powerful tool to improve the nowcast/monitoring of these variables. However, on the revenue side the results are not as consistent although there are significant benefits from the disaggregated approach for many countries.

References

[1] Andreou, E., E. Ghysels and E. Kourtellos (2010): "Regression Models with Mixed Sampling Frequencies", Journal of Econometrics, vol. 158, pp. 246-261.

[2] Artis, M. and M. Marcellino (2001): "Fiscal forecasting: The track record of the IMF, OECD and EC", Econometrics Journal, vol. 4, pp. 20-36.

[3] Baffigi, A., R. Golinelli and G. Parigi (2004): "Real-time GDP forecasting in the euro area", Journal of Forecasting, vol. 20 (3), pp. 447-460.

[4] Bai, J., E. Ghysels and J. Wright (2009): "State space models and midas regressions", Working paper, NY Fed, UNC and John Hopkins.

[5] Clements, P. M. and A. B. Galvao (2008): "Macroeconomics forecasting with mixed frequency data: Forecasting US output growth", Journal of Business and Economic Statistics, vol. 26, pp. 546-554.

[6] Clements, P. M. and A. B. Galvao (2009): "Forecasting US output growth using leading indicators: An appraisal using MIDAS models", Journal of Applied Econometrics, vol. 24, pp. 1187-1206.

[7] Diebold, F.X. and R.S. Mariano (1995): "Comparing predictive accuracy", Journal of Business & Economic Statistics, vol. 13, pp. 253-263.

[8] Foroni, C., M. Marcellino and C. Schumacher (2011): "U-MIDAS: MIDAS regressions with unrestricted lag polynomials", Discussion Paper Series 1: Economic Studies 2011, 35, Deutsche Bundesbank, Research Centre.

[9] Forsberg, L. and E. Ghysels (2006): "Why do absolute returns predict volatility so well?", Journal of Financial Econometrics, vol. 6, pp. 31-67.

[10] Ghysels, E. and N. Ozkan (2012): "Real-time predictions of U.S. federal government budget: Expenditures, revenues and deficits", Working Paper, University of North Carolina.

[11] Ghysels, E, P. Santa-Clara and R. Valkanov (2004): "The MIDAS Touch: Mixed Data Sampling Regression Models", CIRANO Working Papers 2004s-20, CIRANO.

[12] Ghysels, E, P. Santa-Clara and R. Valkanov (2006): "Predicting volatility: Getting the most of return data sampled at different frequencies", Journal of Econometrics, vol. 131, pp. 59-95.

[13] Ghysels, E. A. Sinko and R. Valkanov (2007): "MIDAS Regressions: Further Results and New Directions", Econometric Reviews, vol. 26, pp. 53-90.

[14] Ghysels, E. and J. Wright (2009): "Forecasting professional forecasters", Journal of Business and Economic Statistics, vol. 27, pp. 504-516.

[15] Götz, B. T., A. Hecq and J.-P. Urbain (2012): "Forecasting mixed frequency time series with ECM-MiDaS", Research Memoranda 012, Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization.

[16] Hughes Hallett A., M. Kuhn and T. Warmedinger (2012): "The gains from early intervention in Europe: Fiscal surveillance and fiscal planning using cash data", European Journal of Government and Economics, Vol 1(1), pp. 44-65.

[17] Keereman, F. (1999): "The track record of the Commission Forecasts", Economic Papers, vol. 137.

[18] Koening, E. F., S. Dolmas and J. Piger (2003): "The Use and Abuse of Real-Time Data in Economic Forecasting", The Review of Economics and Statistics, MIT Press, vol. 85 (3), pp. 618-628.

[19] Kuzin, V., M. Marcellino and C. Schumacher (2011): "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area", International Journal of Forecasting, Elsevier, vol. 27(2), pp. 529-542.

[20] Lütkepohl H. (2010): "Forecasting aggregated time series variables: A survey", Journal of Business Cycle Measurement and Analysis, vol. 2, pp. 37-62.

[21] Marcellino, M., J.H. Stock and M. Watson (2003): "Macroeconomic forecasting in the euro area: country specific versus euro area information", European Economic Review, vol. 47, pp. 1-18.

[22] Onorante, L., D.J. Pedregal, J.J. Pérez and S. Signorini (2010): "The usefulness of intra-annual government cash budgetary data for fiscal forecasting in the euro area", Journal of Policy Modeling, vol. 32, pp. 98-119.

[23] Paredes, J., D.J. Pedregal and J.J. Pérez (2009): "A quarterly fiscal database for the euro area based on intra-annual fiscal information", ECB Working Paper Series, No. 1132.

[24] Pedregal, D.J. and J.J. Pérez (2010): "Should quarterly government finance statistics be used for fiscal surveillance in Europe?", International Journal of Forecasting, vol. 26, pp. 794-807.

[25] Perevalov, N. and P. Maier (2010): "On the advantages of dissagregated data: Insights from forecasting the US economy in a data-rich environment", Bank of Canada Working Paper, 2010-10.

[26] Pérez, J.J. (2007): "Leading indicators for euro area government deficits", International Journal of Forecasting, vol. 23, pp. 259-275.

[27] Stock, J.H. and M.W. Watson (2004): "Combination forecasts of output growth in a seven-country data set", Journal of Forecasting, vol. 23, pp. 405-430.

[28] Timmermann, A. (2006): "Forecast combinations", G. Elliot, C. Granger and A. Timmermann: Handbook of Economic Forecasting, vol. 1, pp. 136-196.

9 Appendix

9.1 Boxplots

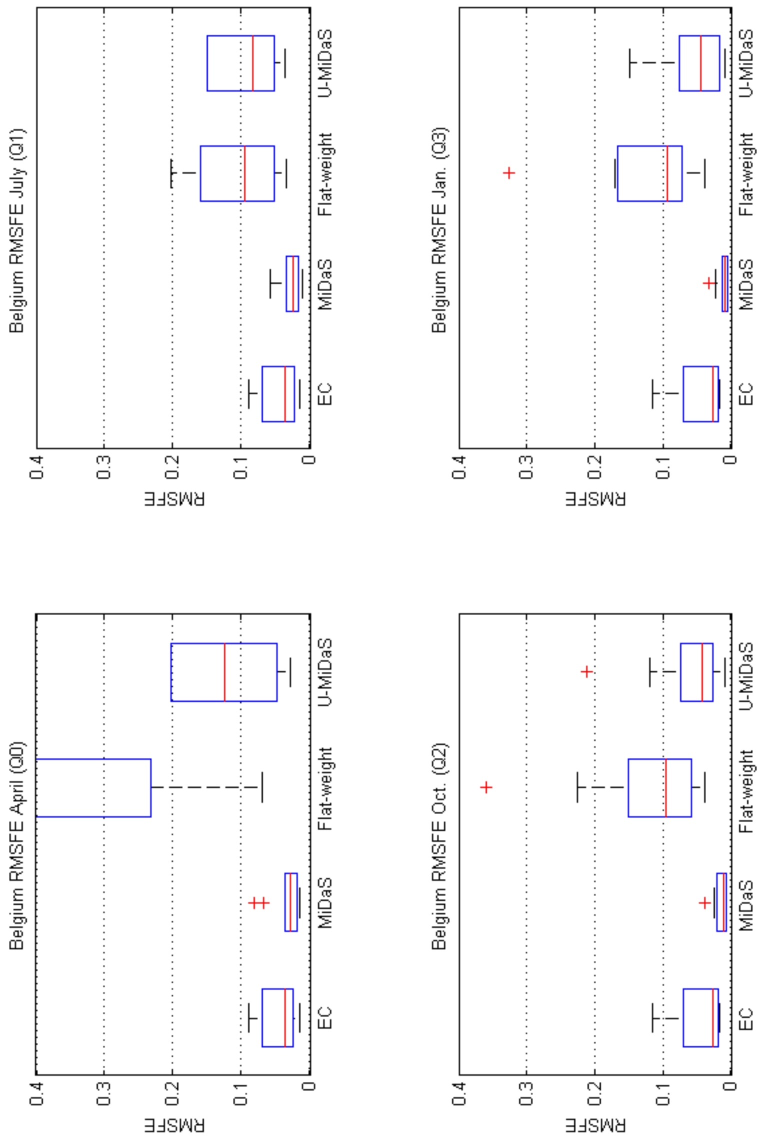


Figure 1: Boxplot-Belgium

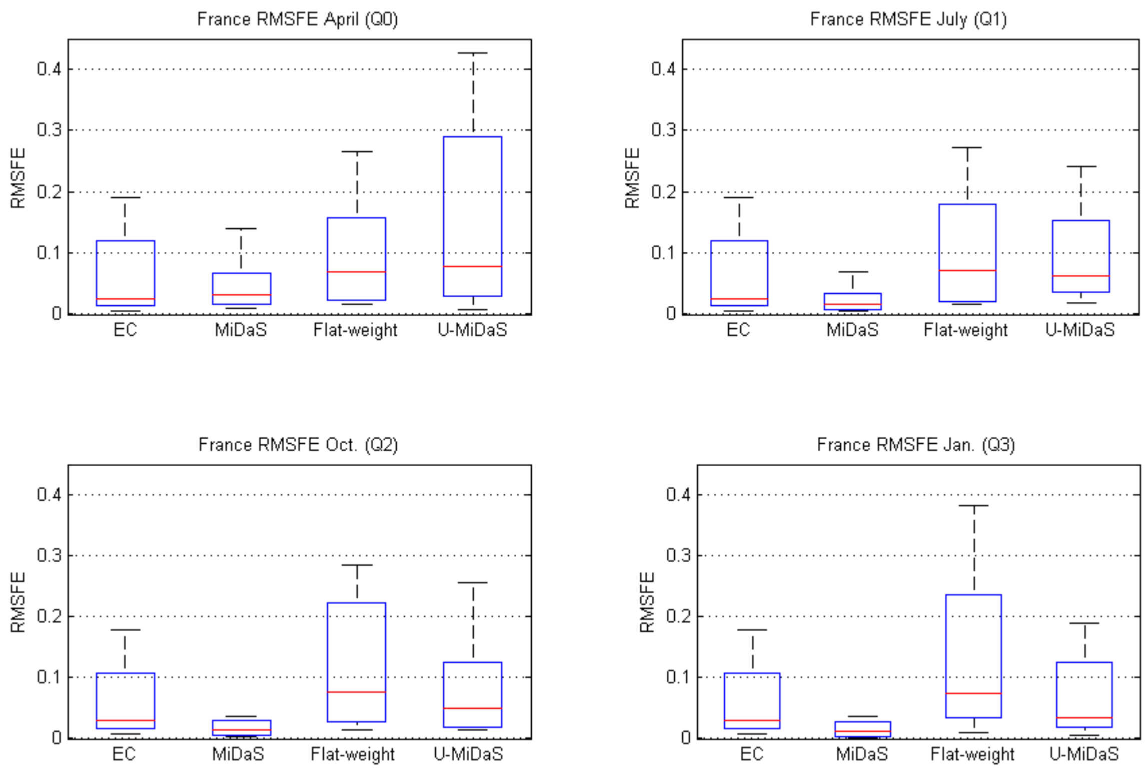


Figure 2: Boxplot-France

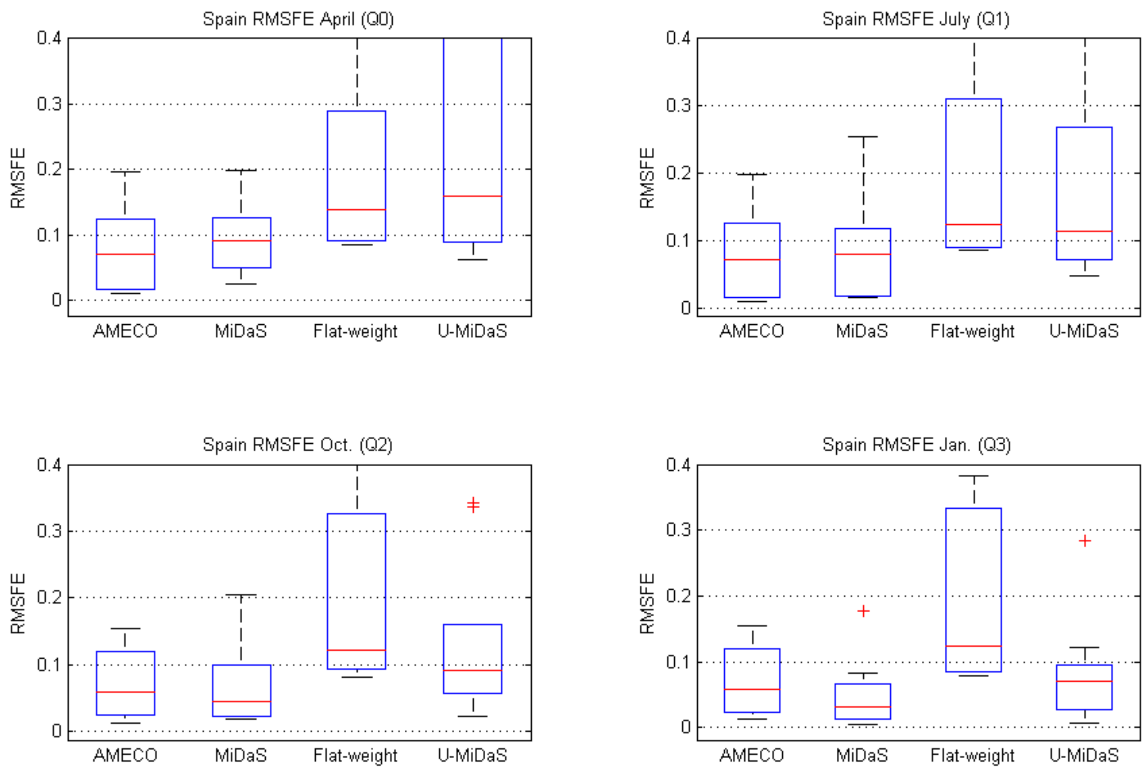


Figure 3: Boxplot-Spain

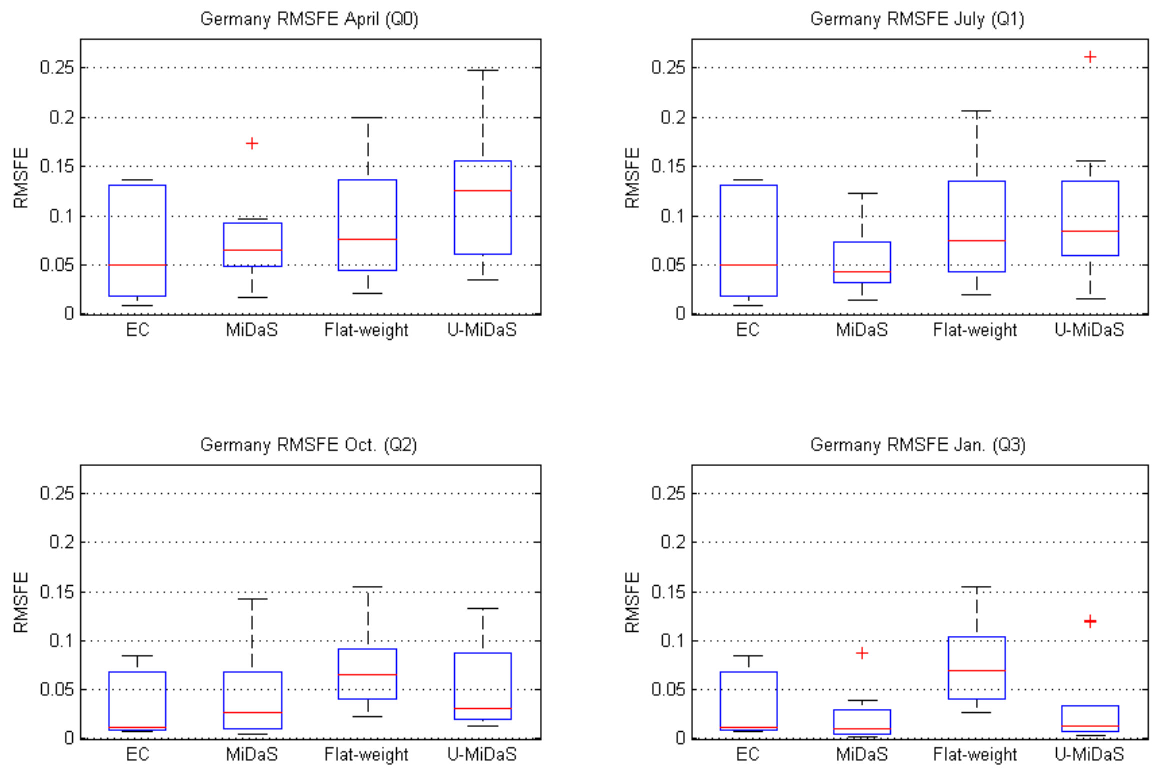


Figure 4: Boxplot-Germany

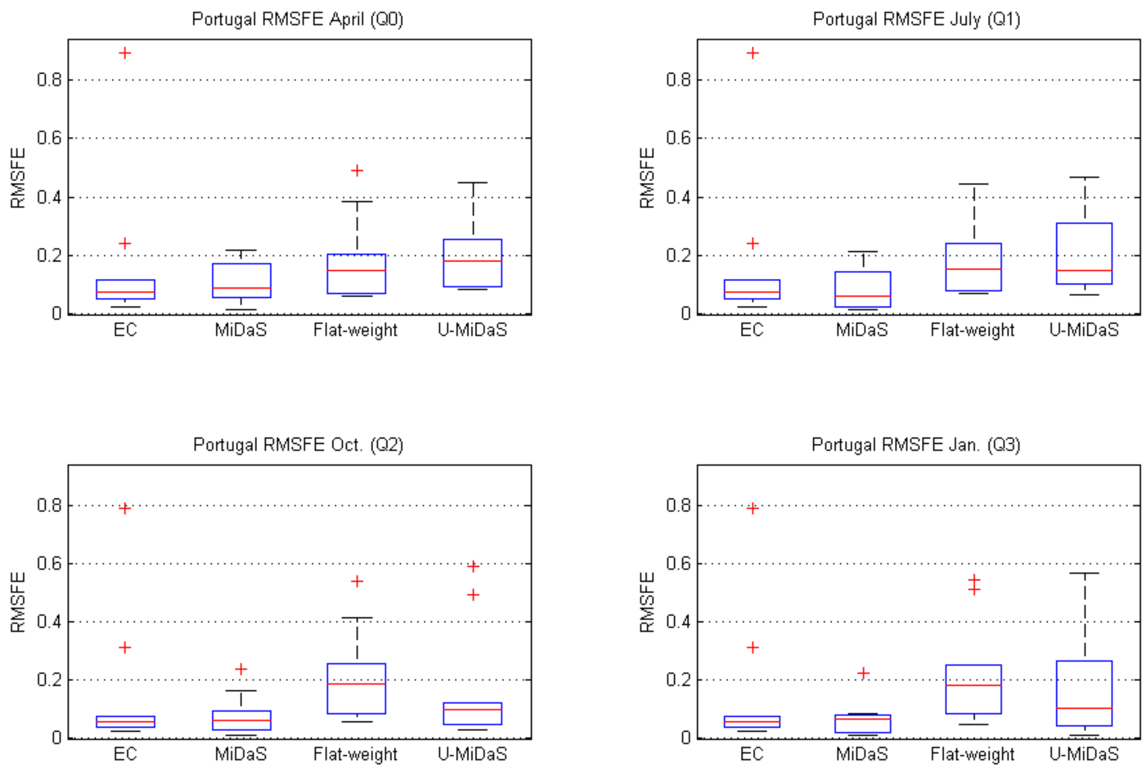


Figure 5: Boxplot-Portugal

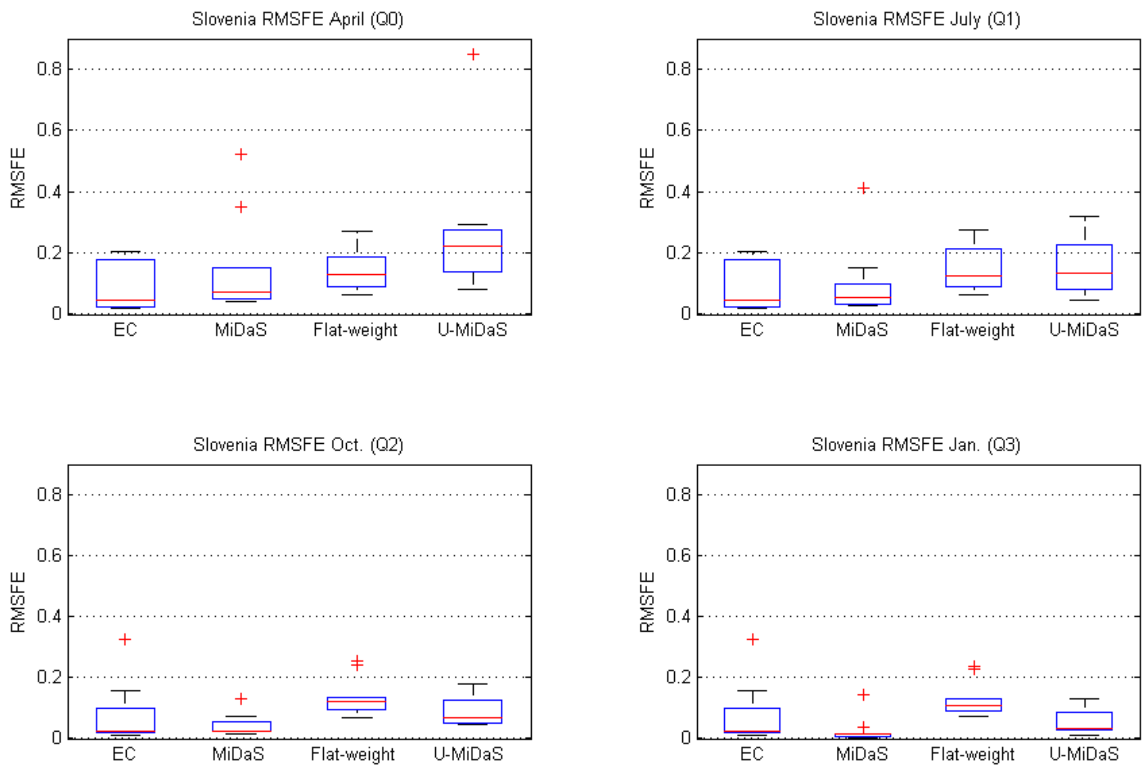


Figure 6: Boxplot-Slovenia

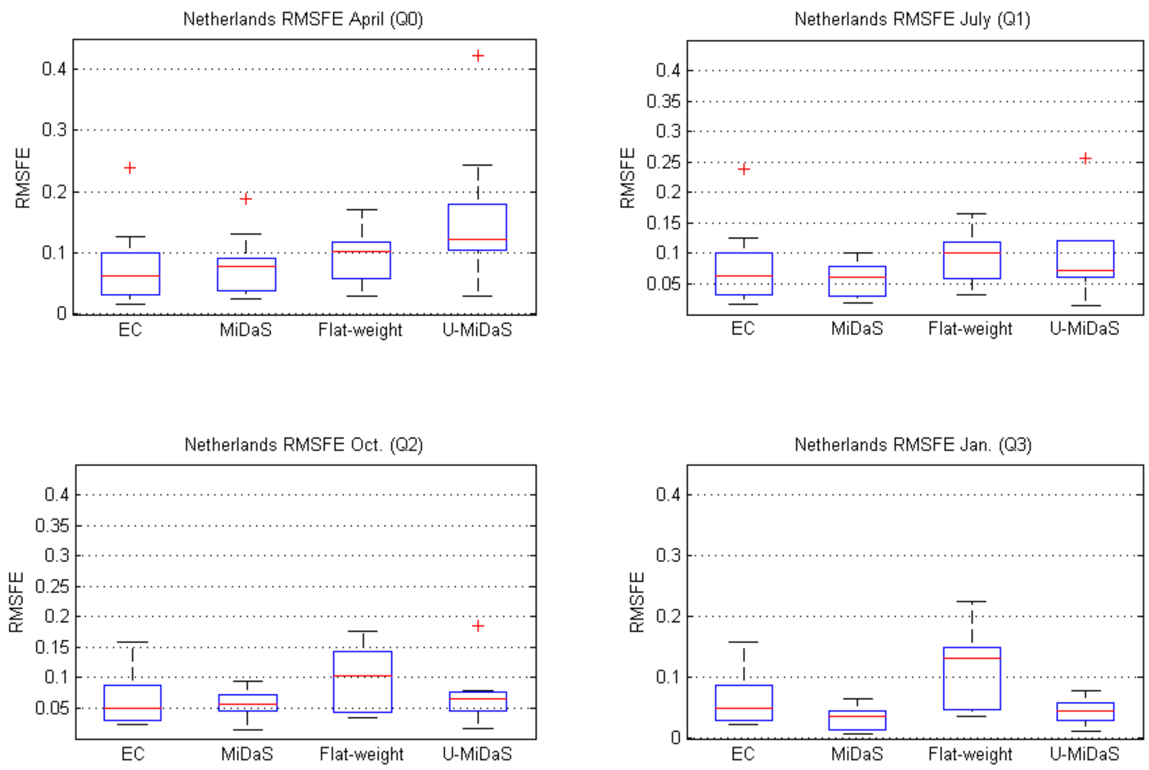


Figure 7: Boxplot-Netherlands

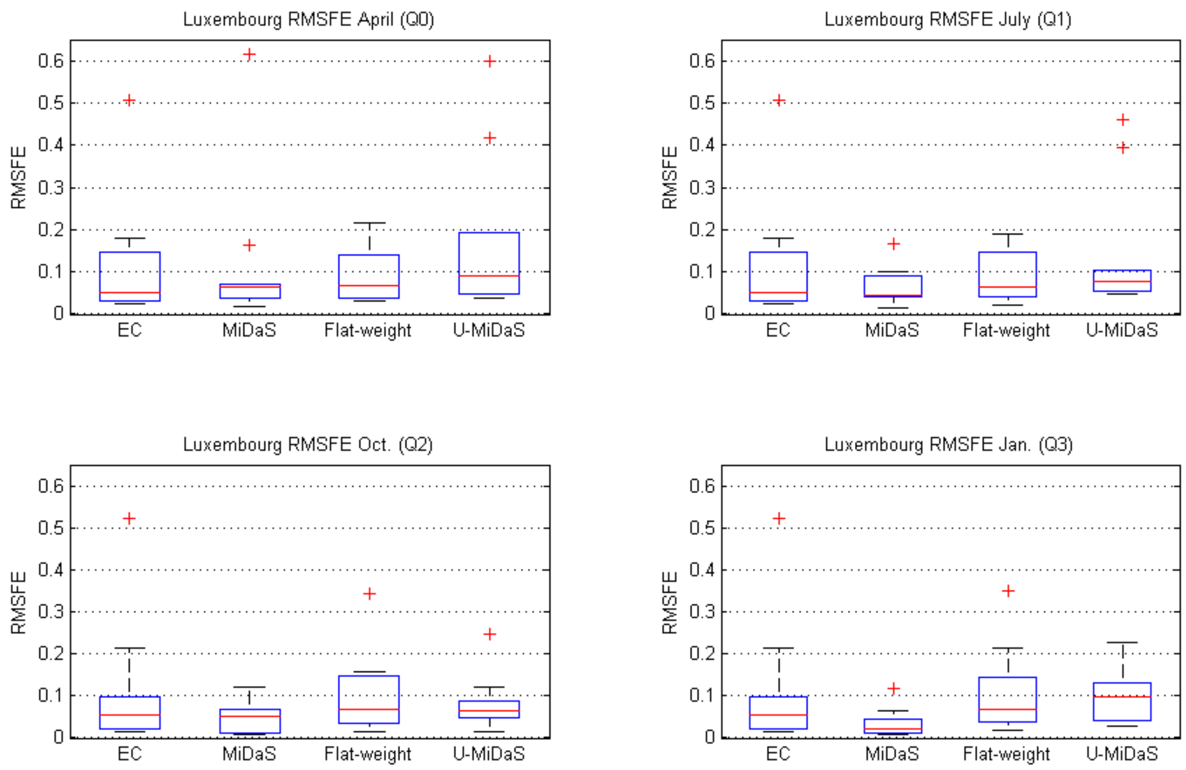


Figure 8: Boxplot-Luxembourg

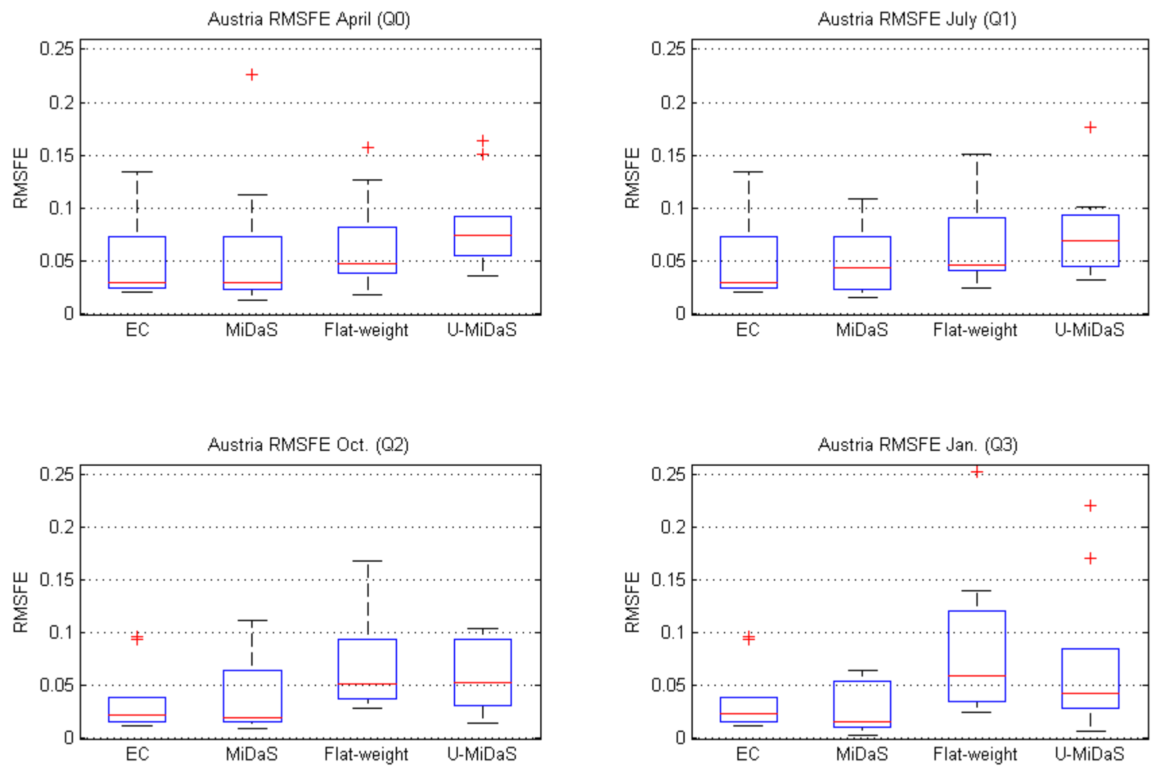


Figure 9: Boxplot-Austria

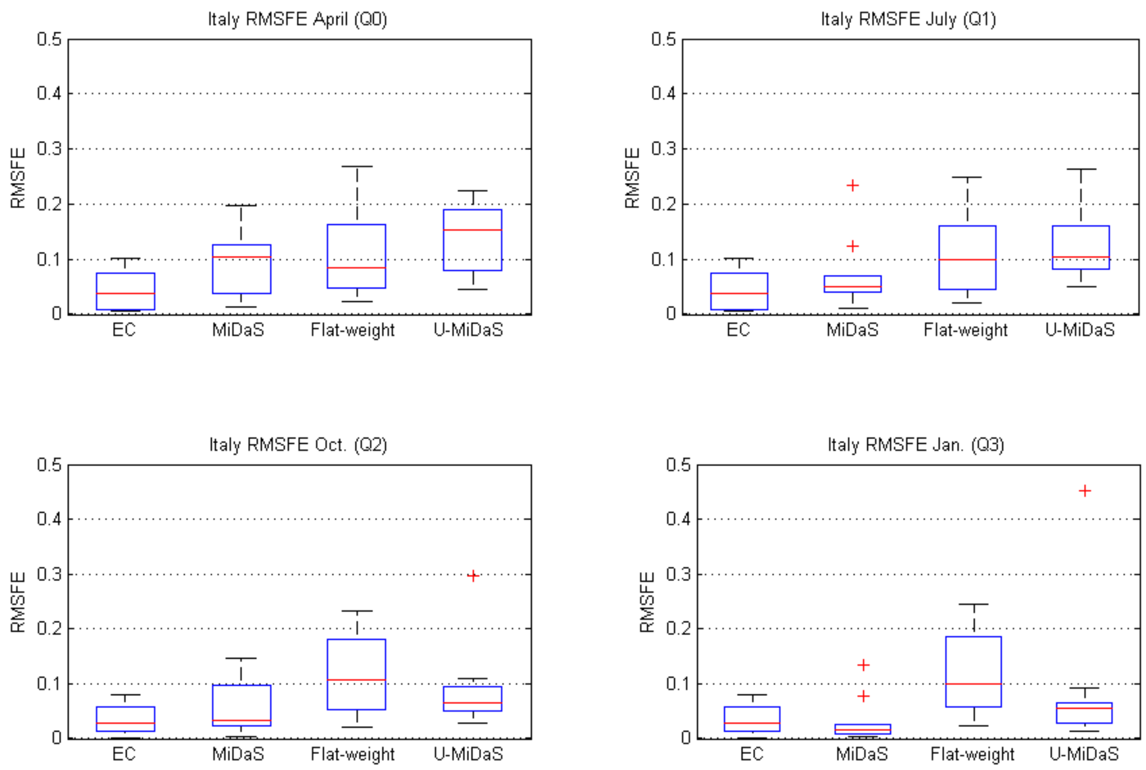


Figure 10: Boxplot-Italy

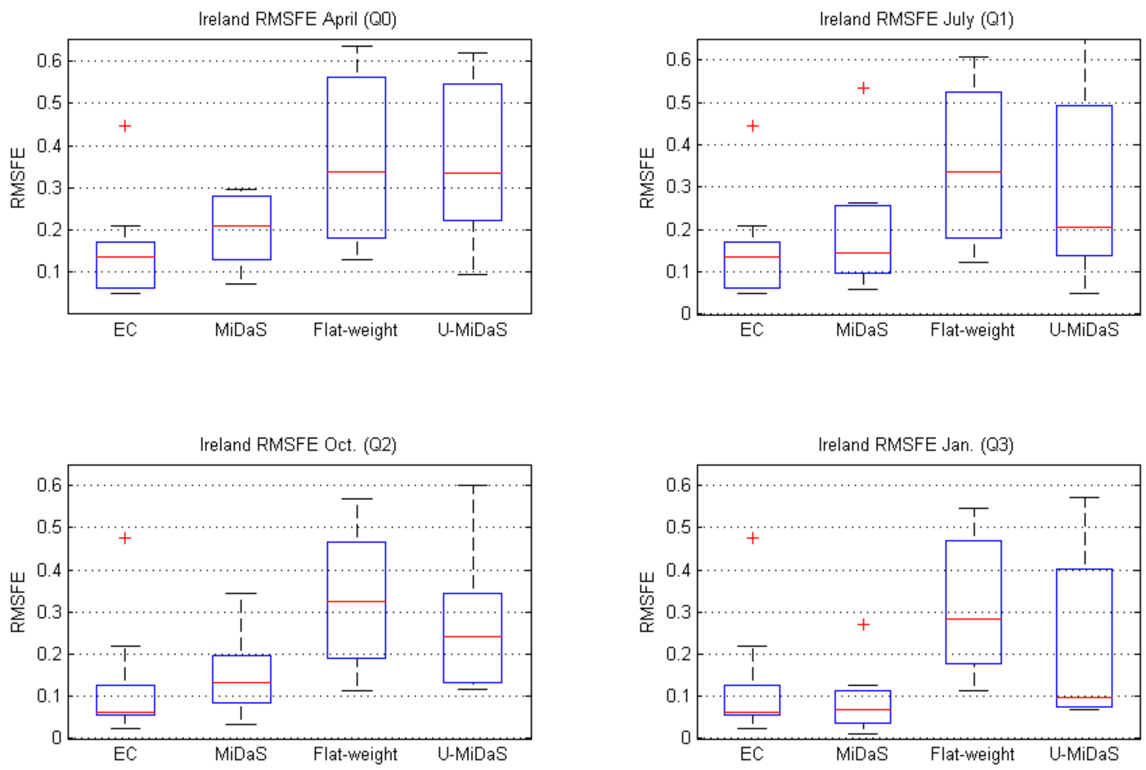


Figure 11: Boxplot-Ireland

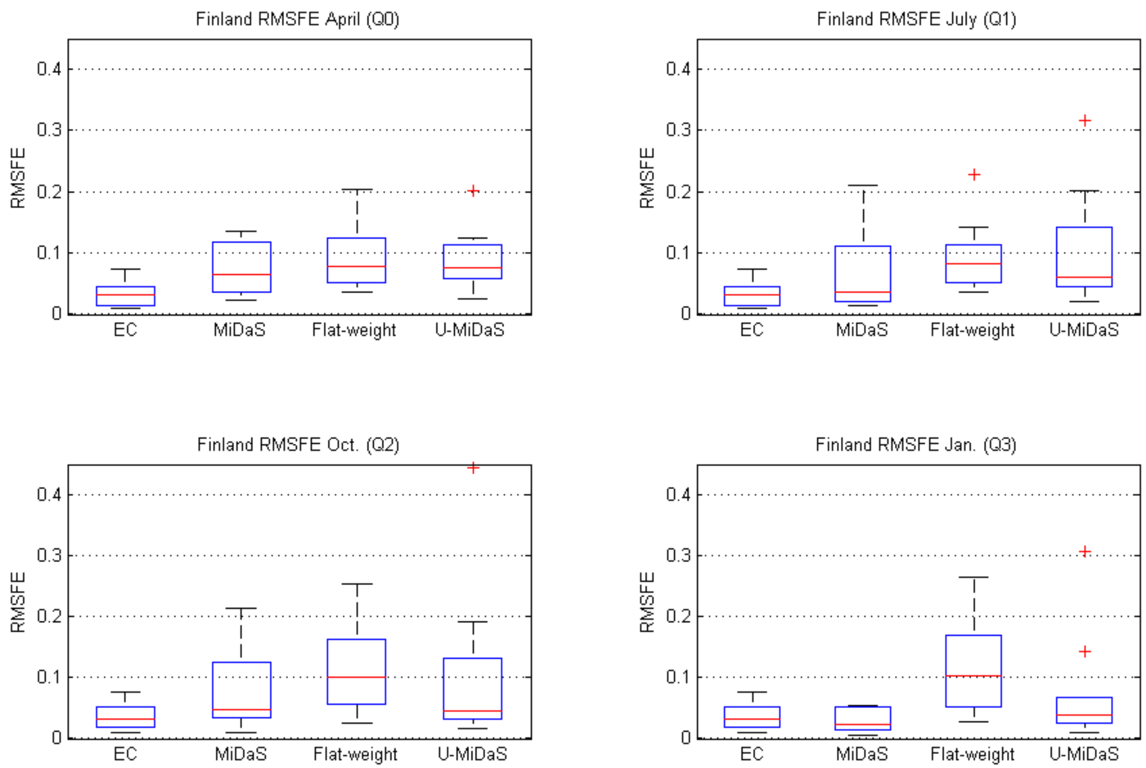


Figure 12: Boxplot-Finland

9.2 Figures of country-specific analysis

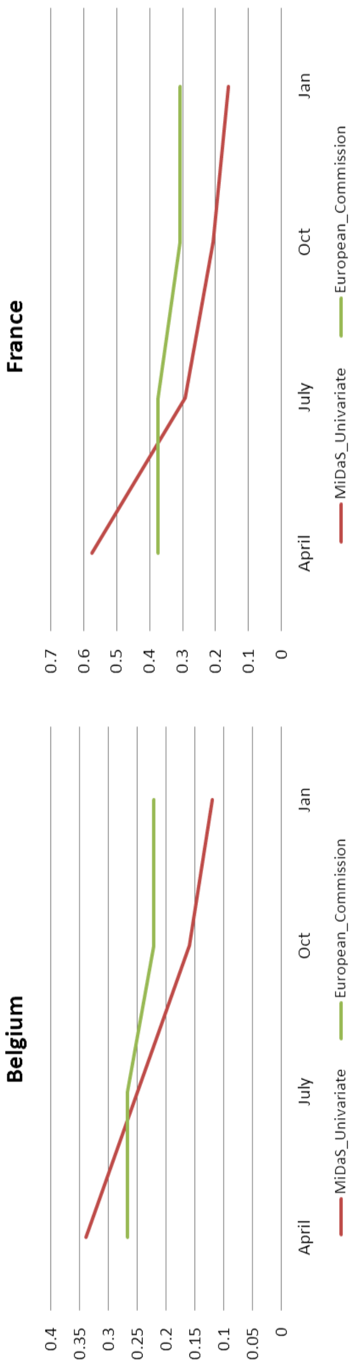


Figure.13

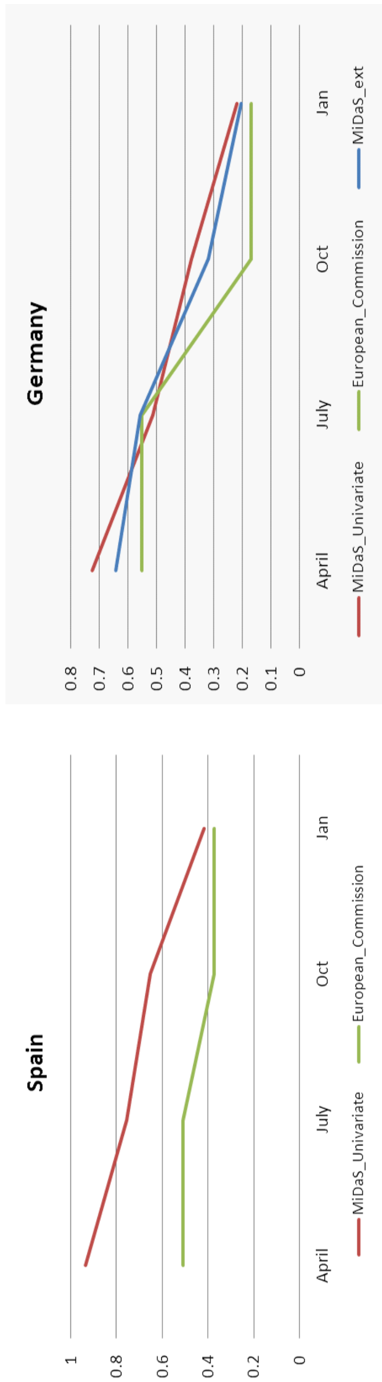


Figure.14

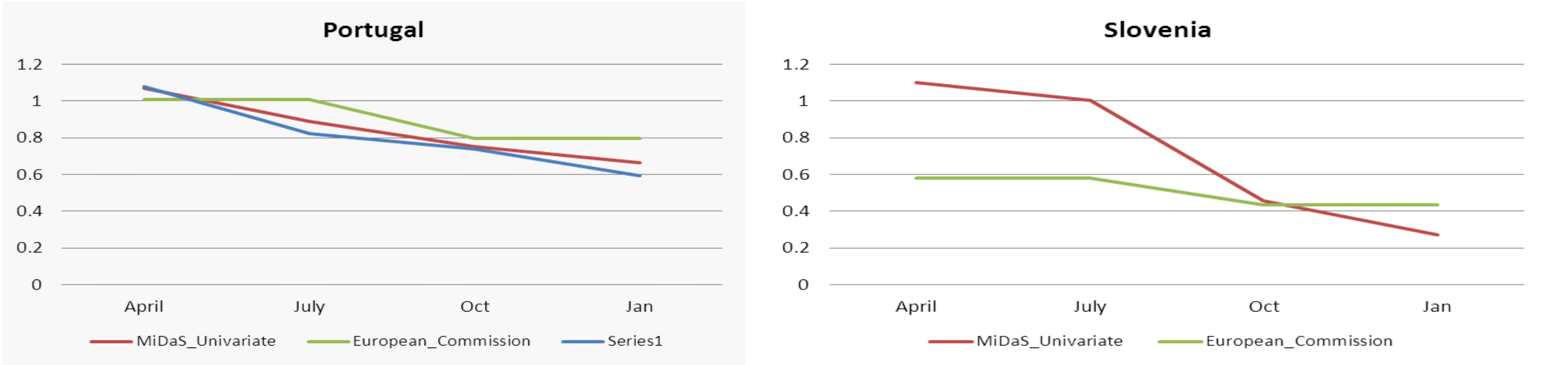


Figure.15

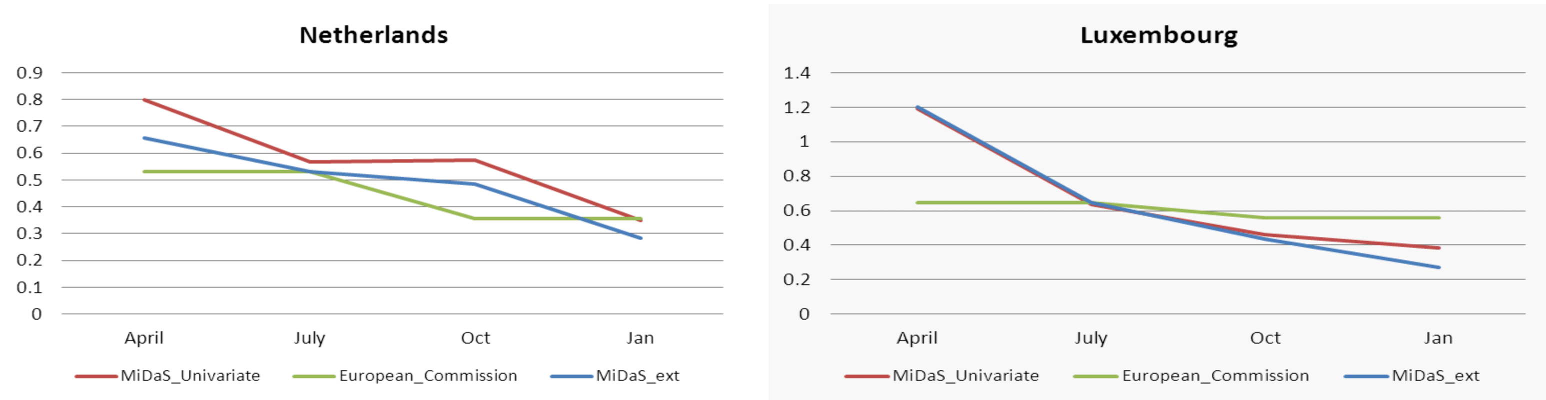


Figure.16

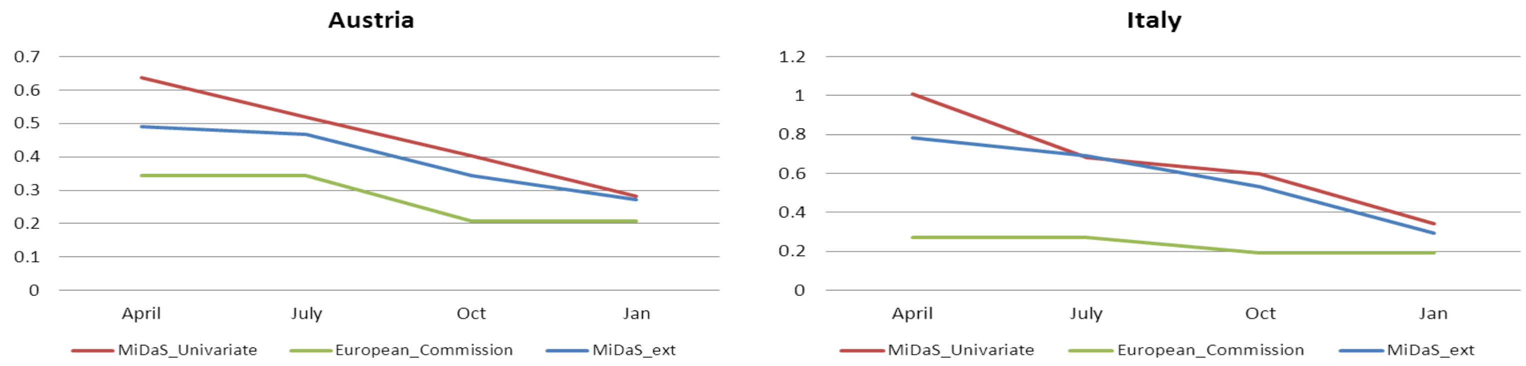


Figure.17

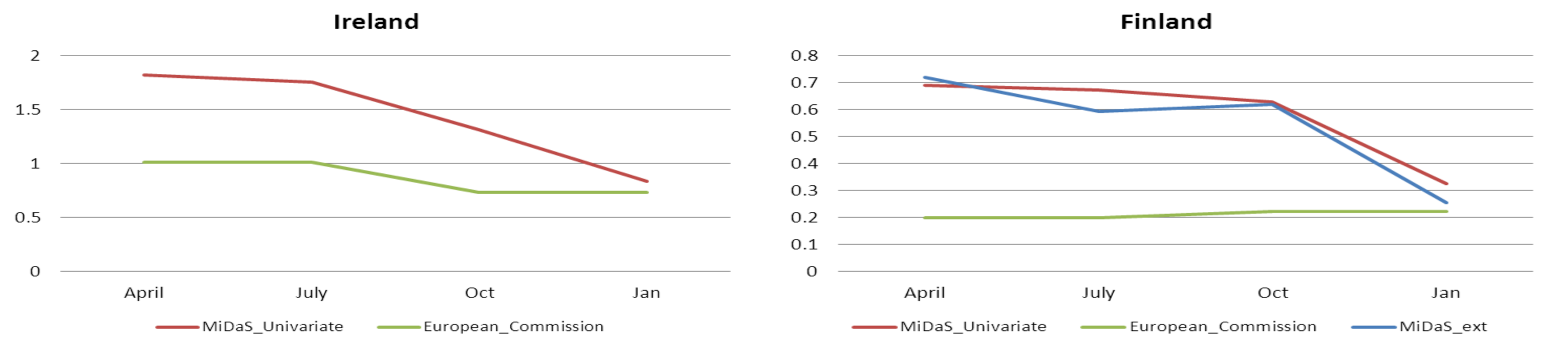


Figure.18

9.3 Figures from aggregated vs disaggregated forecast

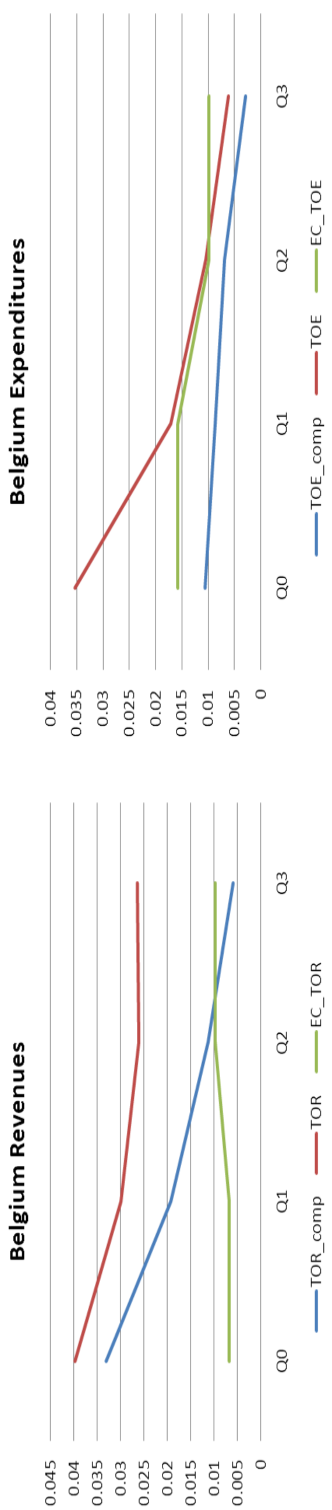


Figure.19

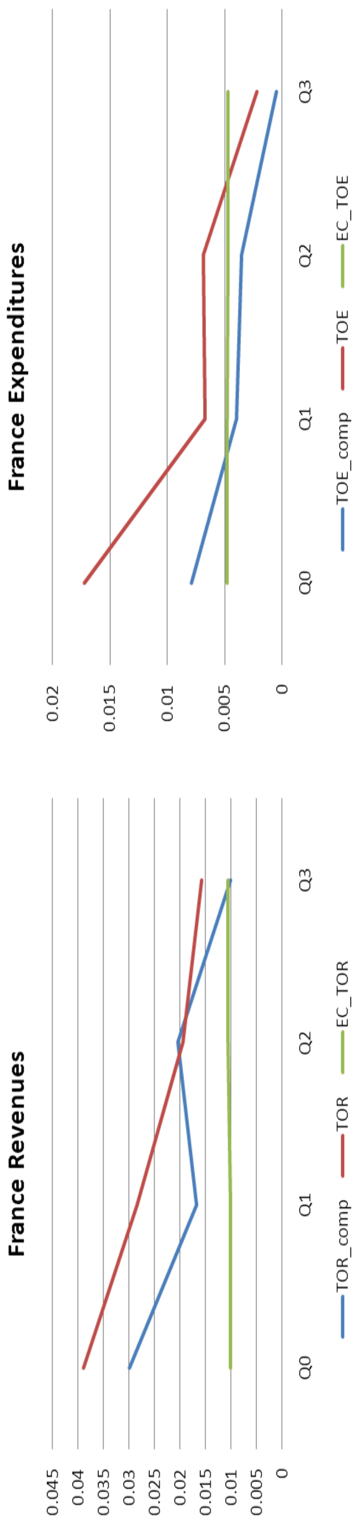


Figure.20

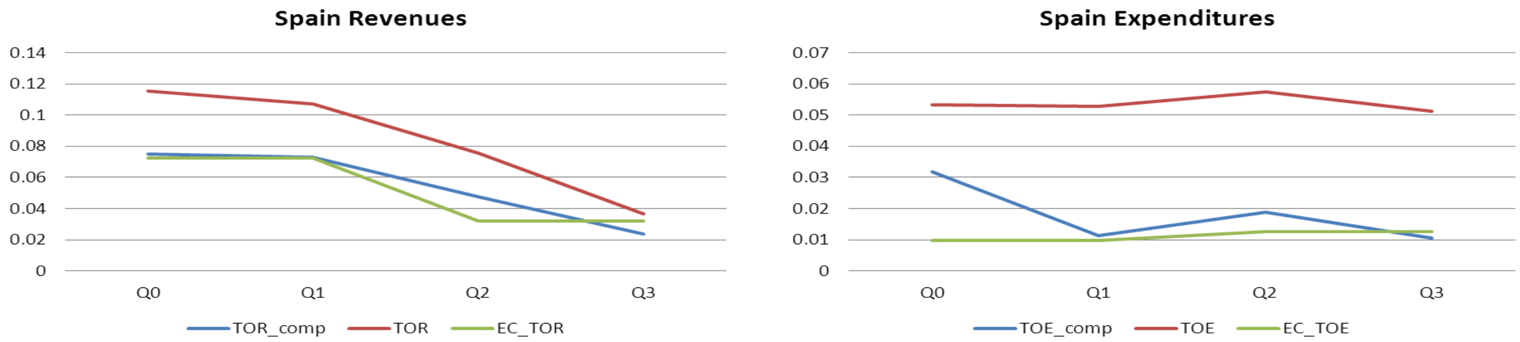


Figure.21

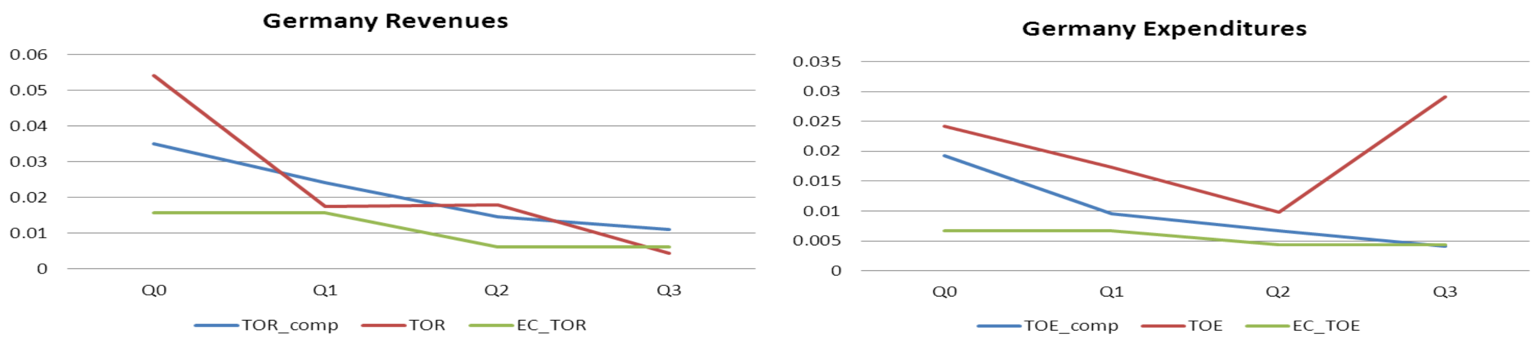
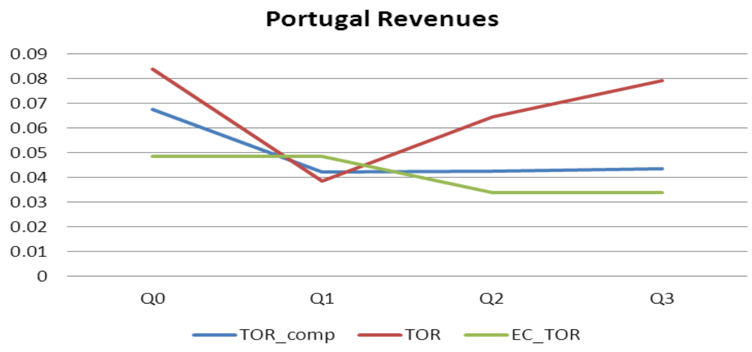


Figure.22



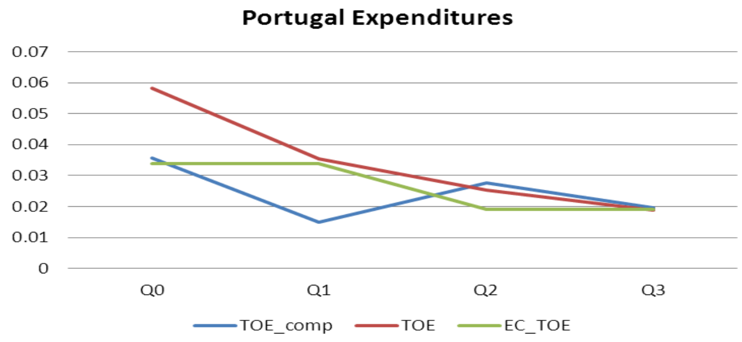
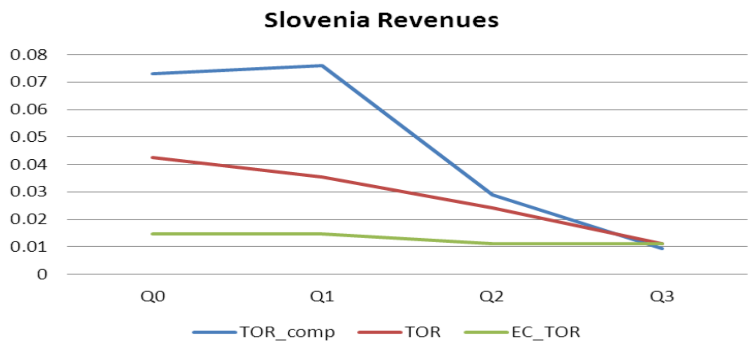


Figure.23



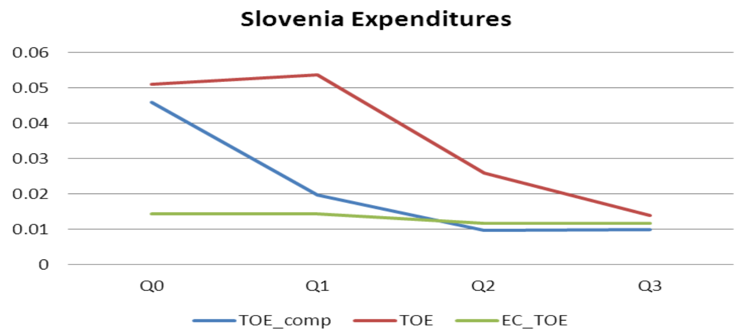


Figure.24

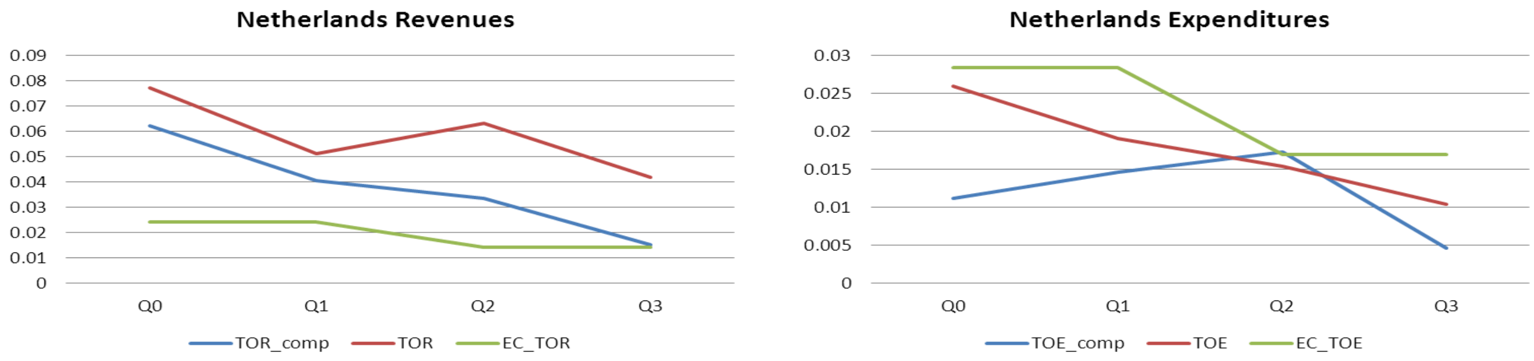


Figure.25

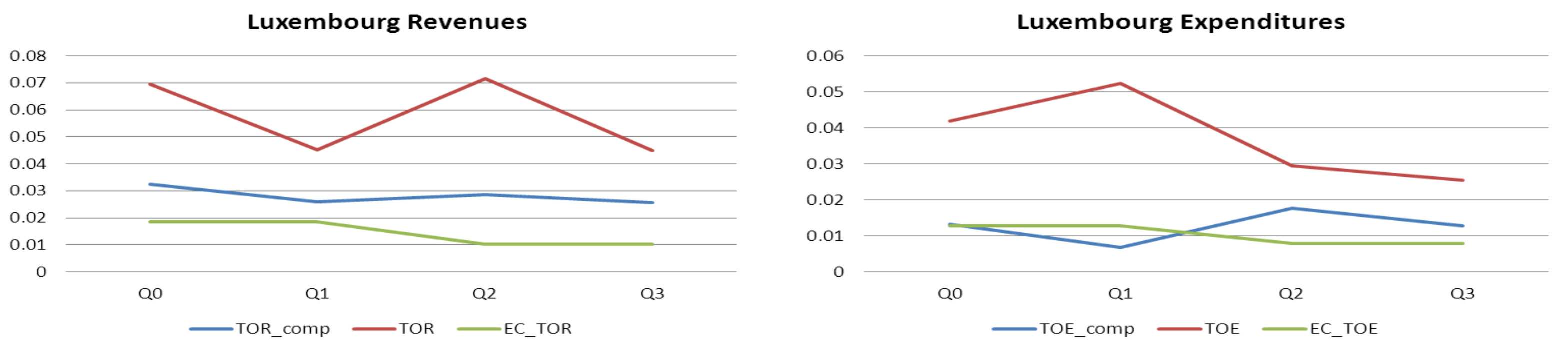


Figure.26

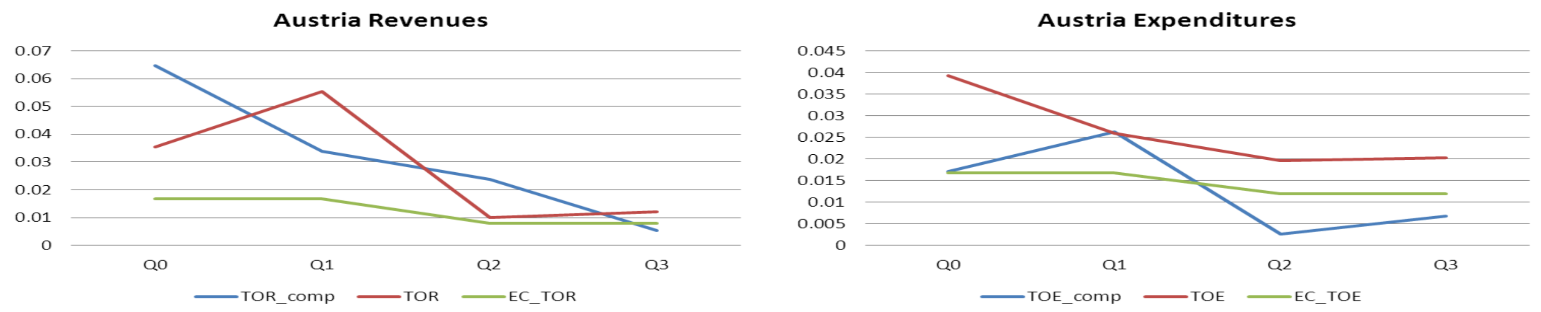


Figure.27

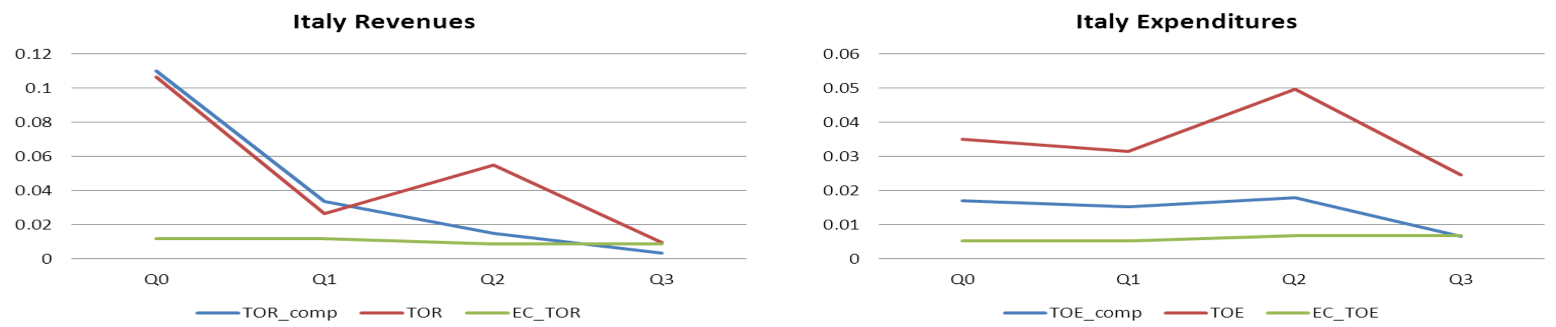


Figure.28

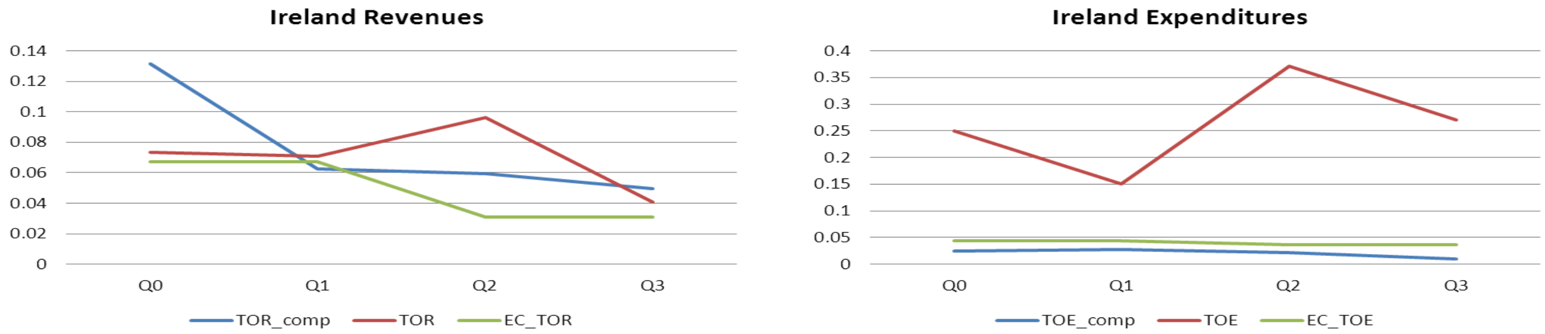


Figure.29

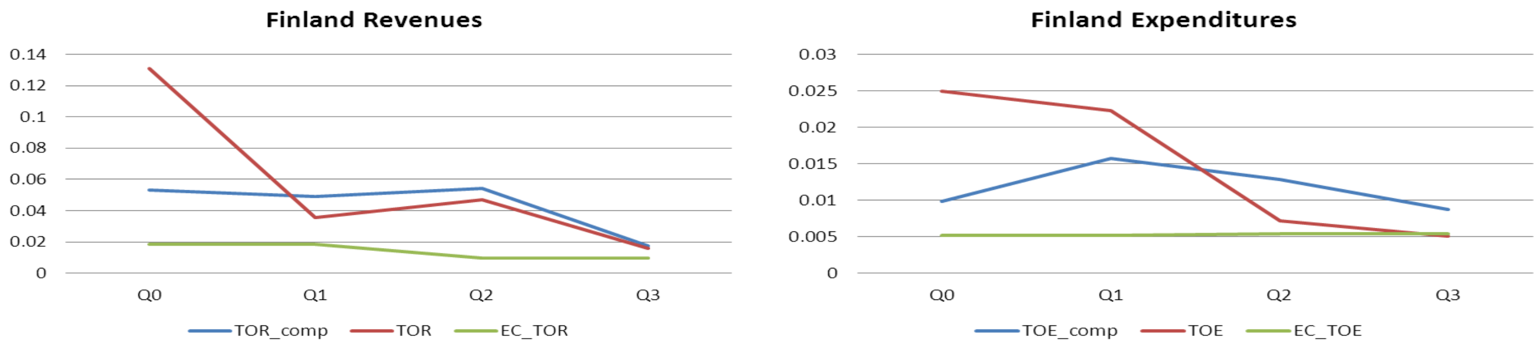


Figure.30

9.4 Diebold-Mariano test

The following tables include the results from the Diebold-Mariano test comparing the forecasting accuracy of the various models. Diebold-Mariano (1995) proposed a test statistic that has as a null hypothesis the equal forecasting accuracy of two models. The root mean squared forecast error (RMSFE) is used as the loss function for testing the statistical difference of the forecast accuracy between models. The RMSFEs used for the tests are the ones obtained in previous sections for each variable and country. Note that the Diebold-Mariano test will assess the mean of the difference between the RMSFEs.

Since we compare always MiDaS with all the other models, when the resulted number is negative it means that MiDaS has lower RMSFE on average. If the negative number reported is also statistically significant, it means that the difference as it appears in the RMSFE, as reported earlier in the boxplots and the graphs, is also statistically significant.[22]

Therefore, in the following tables it is indicated the per country and per quarter performance of MiDaS against the alternative models. The results confirm the basic results mentioned earlier. MiDaS approach outperforms the flat weighting and U-MiDaS approaches in every country and in every quarter with few exemptions.

Focussing on the MiDaS and EC comparison, the sign tends to be negative towards the end of the year when more quarterly fiscal data releases become available (as expected from the boxplots and the graphs presented earlier). However, the statistical significance of the results is mixed across quarters and countries. This indicates that even by taking only the high frequency fiscal data into account when forecasting the annual fiscal data the resulted forecast will not statistically differ from the forecast reported from the EC.

This is a very strong indicator that the quarterly data include very valuable information and should be taken into account when a forecast of fiscal

[22] The statistical significance is noted as:

* for rejection of the null hypothesis at the 10% level of significance.

** for rejection of the null hypothesis at the 5% level of significance.

*** for rejection of the null hypothesis at the 1% level of significance.

When the null hypothesis is rejected then the accuracy of the forecast is statistically significant and the superior model is model one if the sign is negative (model 2 if the sign is positive).

In this case study model 1 is always MiDaS.

data is being performed.

Belgium D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	0.9	-4.1***	-3.9***
Q_1	-0.8	-4.7***	-4.2***
Q_2	-1.6	-7.1***	-3.8***
Q_3	-2.2*	-4.1***	-4.5***

France D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	1.4	-4.0***	-4.8***
Q_1	-1.5	-4.5***	-4.1***
Q_2	-1.8	-4.4***	-3.9***
Q_3	-2.2*	-4.8***	-4.4***

Spain D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	0.9	-6.3***	-6.6***
Q_1	1.1	-5.2***	-5.3***
Q_2	1.0	-5.2***	-4.7***
Q_3	0.1	-5.8***	-3.7***

Germany D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	1.4	-3.2**	-3.4**
Q_1	-1.0	-3.7***	-3.6***
Q_2	2.1*	-2.9**	-1.9*
Q_3	0.2	-5.5***	-1.3

Portugal D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	0.5	-2.0*	-2.9**
Q_1	-1.1	-3.2**	-3.5***
Q_2	-0.9	-3.8***	-3.3**
Q_3	-1.2	-3.6***	-3.5***

Slovenia D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	1.8	0.1	-3.1**
Q_1	0.9	-2.3**	-3.2**
Q_2	0.2	-4.1***	-3.9***
Q_3	-0.8	-4.5***	-3.8***

Netherlands D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	0.5	-1.2	-2.9**
Q_1	0.1	-2.3**	-2.5**
Q_2	1.1	-3.5***	-1.3
Q_3	-0.3	-4.2***	-0.9

Luxembourg D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	0.9	-1.6	-1.9*
Q_1	-0.8	-1.2	-2.9**
Q_2	-1.0	-2.9**	-1.8*
Q_3	-1.3	-3.1**	-2.2**

Austria D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	1.1	-1.7	-1.9*
Q_1	0.8	-1.9*	-2.1*
Q_2	2.1*	-2.7**	-2.3**
Q_3	0.8	-4.5***	-3.4***

Italy D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	2.9**	-0.8	-1.1
Q_1	1.7	-1.9*	-2.0*
Q_2	2.0*	-2.6**	-2.4**
Q_3	1.1	-2.9**	-2.8**

Ireland D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	1.2	-3.8***	-3.8***
Q_1	0.9	-2.5**	-2.4**
Q_2	0.7	-3.1**	-2.9**
Q_3	0.2	-3.9***	-3.1**

Finland D-M Test

MiDaS vs	EC	Flat Weight	U-MiDaS
Q_0	2.8**	-1.0	-0.9
Q_1	2.6**	-1.8	-2.0*
Q_2	2.1*	-2.0*	-1.7
Q_3	0.3	-3.8***	-2.3**

9.4.1 Testing the accuracy of the forecast after including an autoregressive term in MiDaS regression

It has been mentioned in the main text that the inclusion of an autoregressive (AR) term can cause multicollinearity issues and for that reason such a term has not been included throughout the analysis so far. However, since the current work is not focused on the specific estimates of the regressions but only on nowcasting the dependent variable, an AR term is included for comparison purposes. Specifically, the same approach is being followed for all three regressions but now one lag of the dependent variable is always included for the estimation of the same fiscal variables.

Initially, it is found that ADL-MiDaS (MiDaS with an AR term), as it is outlined in equation (4), is the best approach compared to the alternatives of flat-weight and U-MiDaS regressions, both with an AR term.[23] That means that the RMSFEs of the ADL-MiDaS are significantly lower than the other two approaches. This result is in accordance with the results presented in the previous subsection without the inclusion of an AR term.

As a last step, the RMSFEs of DL-MiDaS (MiDaS without an AR term), as in equation (3), and the ADL-MiDaS are compared. Using the D-M test the following results are obtained:

	D-M test
Belgium	0.72
France	1.65
Spain	3.01**
Germany	1.12
Portugal	-0.37
Slovenia	1.98*
Netherlands	0.82
Luxembourg	1.04
Austria	-0.10
Italy	0.60
Ireland	1.37
Finland	0.24

The positive sign of the D-M test means that DL-MiDaS regression, without an AR term, has lower RMSFEs than the ADL-MiDaS. Thus, the results suggest that MiDaS approach without an AR term is superior to the ADL-MiDaS, due to the positive sign for most of the results. However, the

[23]The results of the forecasting accuracy comparison between the three different models with an AR component are not reported for space purposes and because they are qualitatively identical with the results without an AR component presented earlier.

difference is statistically significant only for Spain and Slovenia. That means that qualitatively the forecast accuracy among the two approaches is similar on average, although the DL-MiDaS reports lower RMSFE for most of the countries.

The reason for obtaining a better forecast with the MiDaS approach without the AR term is due to the information included in the AR term. Specifically, the AR term includes the annual information of the dependent variable from the previous period. But this information is already included in the DL-MiDaS approach because the quarterly data of that year are always included. Note that the quarterly data enter MiDaS with at least 6 lags, which is translated into one and a half year of information of the dependent variable. This makes the information of the AR term redundant. Therefore, the AR term cannot improve the accuracy of the forecast in the current analysis because it does not add any new information to the model.

9.4.2 Testing the statistical significance of the additional sample

It has been shown that the additional sample that was available for some variables and for some countries can improve the forecast accuracy. Therefore, the Diebold-Mariano test will be used to test the statistical significance of the improvement. In particular, the same procedure as before will be followed. In this case a negative test statistic means that the additional sample has lower RMSFE and improves the accuracy of the model.

The results indicate that for every country where the additional sample was available the forecast is more accurate in terms of lower RMSFE (since the sign is always negative). However, the improvement is statistically significant only for Portugal, Netherlands and Austria.

D-M test

Germany	-1.02
Portugal	-1.88*
Netherlands	-3.79***
Luxembourg	-1.01
Austria	-2.38**
Italy	-1.65
Finland	-1.28

9.4.3 Testing the statistical significance of aggregated versus disaggregated approach

We finally assess whether the disaggregated approach can result in statistically significant improved results when forecasting total revenues and total expenditures. If the Diebold-Mariano test reports negative test statistics, then the disaggregated approach produces more accurate forecasts than the aggregated approach (lower RMSFE).

The results in the table verify what it has been shown earlier with the graphs. In particular, it is indicated that on the expenditure side there is a consistent improvement from the disaggregated approach for every country (negative test statistic) and also the improvement is statistically significant for most of the countries. However, on the revenue side the results are more mixed but, for at least half of the countries, the disaggregated approach improves the forecast accuracy and is also statistically significant.

D-M test

	Total Revenues	Total Expenditures
Belgium	-4.46***	-2.38**
France	-2.45**	-2.57**
Spain	-4.91***	-7.51***
Germany	-0.37	-2.33**
Portugal	-2.27**	-1.49
Slovenia	1.73	-2.34**
Netherlands	-4.57***	-1.68
Luxembourg	-4.89***	-3.07**
Austria	0.33	2.72**
Italy	-0.82	-5.84***
Ireland	0.27	-5.11***
Finland	-0.64	-0.63