Fiscal Forecasting in Italy

Laura Carabotta

Ph.D. in Economics

Thesis title:

Fiscal forecasting in Italy

Ph.D. Student:

Laura Carabotta

Advisors:

Elisenda Paluzie i Hernàndez

Peter Claeys

When you set out on your journey to Ithaca,
pray that the road is long,
full of adventure, full of knowledge.
The Lestrygonians and the Cyclops,
the angry Poseidon -- do not fear them:
You will never find such as these on your path,
if your thoughts remain lofty, if a fine
emotion touches your spirit and your body.
The Lestrygonians and the Cyclops,
the fierce Poseidon you will never encounter,
if you do not carry them within your soul,
if your soul does not set them up before you.

Pray that the road is long.
That the summer mornings are many, when,
with such pleasure, with such joy
you will enter ports seen for the first time;
stop at Phoenician markets,
and purchase fine merchandise,
mother-of-pearl and coral, amber and ebony,
and sensual perfumes of all kinds,
as many sensual perfumes as you can;
visit many Egyptian cities,
to learn and learn from scholars.

Always keep Ithaca in your mind.
To arrive there is your ultimate goal.
But do not hurry the voyage at all.
It is better to let it last for many years;
and to anchor at the island when you are old,
rich with all you have gained on the way,
not expecting that Ithaca will offer you riches.

Ithaca has given you the beautiful voyage.
Without her you would have never set out on the road.
She has nothing more to give you.

And if you find her poor, Ithaca has not deceived you.
Wise as you have become, with so much experience,
you must already have understood what Ithacas mean.

Constantine P. Cavafy (1911)

Acknowledgements

During these long years through my doctoral thesis, my journey has been characterized by extremes. I have had peaks of excitement and happiness: The joy of finding new ideas, of being inspired by the research, of the light bulb that lights up when you find the solution you've been searching for, and the pleasure of the same pattern of travel. And I also have had moments of panic! Moments of existential emptiness, of decentralization, of feeling immensely tired, of disappointment and of uncertainty not knowing where to make my next step. I am sure that it is only thanks to the support and confidence of the people who have followed this trip with me, that I could balance these forces, finding the courage to take that step each day and then enjoy it with them.

I would like to thank first and foremost my supervisors Elisenda Paluzie i Hernàndez and Peter Claeys for the time they spent with me. First of all, thanks to Elisenda for the support, the trust and the possibility that she gave me to make my dream of a Ph.D come true. Peter for the work done together, for the insights, the reflections he ispired me and for the mechanisms and research questions that he taught me.

My most affectionate thanks go to Raúl Ramos, friend and teacher, who supported me and helped me every time I needed it. I am infinitely grateful to him for the time he dedicated to me for academic help and for moral support. I'm really glad to have the opportunity of being able to learn so much from him. This work is deeply, happily influenced by his support. He will always have my friendship, thank you again.

I thank my dear friend Alessia Matano. Her presence in my life in Barcelona and during my doctoral journey was fundamental. From being close to her, I learned what it means to be a researcher and not just simply be doing research. From her I learned to look critically at the facts, to remain lucid in the face of reality. I learned the importance of depth, of curiosity, of the ability to look beyond the questions and of taking the overall view. I am thankful to her for all the hours of intensive conversation, for supporting me in the more difficult choices, for her affection, and especially for being who she is - as a researcher and as a beautiful person.

My kind thanks go to Odyscea Kian, for her professionalism in the English revision of this work and for her availability and sweetness. She was close to me during my Ph.D. journey and she gave me a great deal of security for the work we did together. She has been a guardian angel, a reliable, welcoming presence and, above all, a great friend. To her, I offer all my gratitude.

I thank the academic staff of the Faculty of Economics of the University of Barcelona. In particular, Josep Lluís Carrion-i-Silvestre for econometric discussions and for his kindness and sympathy. Aurelio Tobias, because his lessons helped immensely to facilitate Stata programming of my calculations. I thank him for his very kind manner and support, and for his profoundness as a professor and as a person. Thanks to Luis Díaz Serrano, my office mate and friend, for the conversations about life and the economy, for his mature reflections, for offering me his critical thinking ability and his friendship, which I warmly reciprocated. Thanks to Kristian Estevez for the academic discussions and for reviewing my English at various moments, especially for revising the Ph.D. project. I also thank him for his friendship and kindness.

Each time that I had a problem as a graduate student, I found efficient solutions with the help of the administrative staff of UB economics. Any doubt or request for help was solved by committed, fantastic people, in particular Jordi Roca. Thanks to Elisabet Viladecans-Marsal, Director of the Ph.D., for the organization of the doctoral program and for her availability. I also thank Bibiana, Susana and M.Angels for their sweetness and sympathy.

This PhD has been achieved thanks to the permission granted to me, as a public servant, by my administration: the Ministry of Economy and Finance, The State General Accounting Department, Inspectorate General for Budget. All my gratitude go to my heads: Biagio Mazzotta, Mauro D'Amico and Liliana Dinaro. I thank them very much for the opportunity they gave to me, for the trust shown in me and the support in this project which is part of a path of personal and professional growth. I warmly thank them for their understanding and aid during this time.

Among the most important moments of the research as a Ph.D. student, I remember with happiness the lectures and conferences in which I participated. Being there among the fathers of the literature and being able talk to them and exchange ideas with them was one of the most exciting things that I experienced. I especially thank Domenico Giannone for the insights on the second and third Chapters, Barbara Rossi for the kind econometric discussion of the third Chapter. I thank the academic staff of the Workshop of Time Series Econometrics in Zaragoza (Spain) for useful debate and insights about the second and third Chapters. Thank to Stefano Fantacone for interesting comments about the second Chapter. Thanks to Tommaso Proietti and Alessandro Giovannelli, for the interesting econometric discussion about the fourth Chapter and to Ottavio Ricchi for the debate on academic and institutional points of view about the thesis as a whole. I would also thank my dear friends Manuela Vecchi, associate professor of physics at the University of Sao Paolo and Emiliano Mancini, researcher of biology at the Terza Università of Rome. Although working in different areas of research, their advice has helped and supported me immensely.

During the years of my doctorate thesis, I found myself alone in a foreign country. Sometimes it was difficult to cope with loneliness, but looking back there were just a few moments, because I was lucky to meet people who have formed part of my family in Barcelona. I would like to thank Jordi, for being my partner with love and patience during the first two, and the most difficult, years of the Ph.D.; Angels, who welcomed me into her home and for her friendship, for our long chats, for the affection that I will never forget. Xantal and Miguel for being my sister and brother in one of the friendliest places I've found, for sharing with me their spirituality and force, which I carry with me every day. Isabel and Cecilia - people who most of all I felt to be similar to me, I thank them for the friendship and artistic topics we shared and the happiness that I had at every meeting. I also thank my friends around the world, Sandy in Ireland and Raùl in Peru. And I also thank all my friends from home who were close to me and have made me happy when they visited me in Barcelona: Giuggio, Cristina, Gisela, Alessia, Roberta, Liana, Marta, Marianna, Adriano, Marcello and also my dear friends and colleagues Silvia and Roberta.

Last but not least, I thank my family - my father and my sister who encouraged me once again during this experience. To them, I offer all my love and my gratitude. And I thank my mother, who is close to me even if from far away. I dedicate this work, with love, to her.

Table of Contents

List of Tables

Table of Figures

Chapter I.

Introduction

I.1. Background and motivation

After the economic and fiscal crises of 2008, deficit forecasts are playing an increasing role in macroeconomic policy decisions. With the Treaty of Maastricht and the European Semester, European Union (EU) Member States are required to provide accurate estimates of the main variables of public finance. The system of accounts, in general, provides that the Government is obliged to explain to the country and to the Parliament the results of its decisions in terms of actions, policies, and consequences (Hameed, 2005, Leal et al., 2008). The Government, then, is subject to both the internal control of the country, and to external control at the European level. The internal laws ensure that they comply with the need for balanced public finance, proper timing of the publication of official documents, and proper legal and administrative functioning. From an external point of view, the European Commission is invested by the Member States with the power to set controls, establish policies, and commit the Member States to respect the Stability and Growth Pact (SGP). The Government provides public finance plans that are consistent with the fiscal policies of the country and in line with the fiscal and economic conditions of the European Commission. However, economic planning, public finance forecasts, and national accounts produce systematic errors that are not adaptive and can bring very serious consequences. An historical case is given by the recent crisis of 2008-2014. For example, between 2010-2011, countries had to spend public resources to support the banking system, especially in those cases where international institutions wanted to avoid sovereign default (like Portugal, Ireland or Greece), through the provision of large loans from the International Monetary Found (IMF) and EU, called the "bailout plan". The aim was to avoid possible defaults, though it resulted in severely restrictive fiscal policies on public accounts (austerity). Such policies brought a reduction of consumption, and a downward spiral in production. The Governments did not have, during this crisis, the right methods to predict the macroeconomic variables in adaptive and flexible ways and make them available in real-time. This lack of proper instruments produced a strong difference between forecasts and the actual value of the main macroeconomic variables. Following the crisis, the European Commission changed its measures with a new framework. By now the link between the qualities of the fiscal forecast frameworks and budgetary discipline is required at the international level.

It is against this framework that the Member States introduced the new control procedures for national budgets. Forecasts started to be considered in the European Council meeting on 1-2 March 2012, in which 25 European leaders signed the Treaty on Stability, Coordination and Governance (TSCG) aimed at strengthening fiscal discipline and introducing stricter surveillance within the euro area. The European Commission now sets up a yearly cycle of economic policy coordination called the European Semester. Each year the European Commission undertakes a detailed analysis of Member States' programmes of economic and structural reforms, and provides them with recommendations for the next 12 to 18 months. With the European Semester, the EU has an ex-ante way of evaluating Member States' plans for sound public finances and can therefore anticipate any deficit issues and make recommendations early enough so that Member States can still adjust their plans and make reforms accordingly. This gives the Governments a new focus on presenting accurate forecasts about public finance variables to the European Union. The deficit forecast, in particular, is the most important economic and political variable in this process of forecast evaluation. Its future evaluation must not exceed 3% of GDP. For countries with very high levels of debt and deficits, this new procedure is supposed to be particularly helpful to avoid deviations from initial budget plans.

This new European approach and its relevance to national and international economic policies has motivated new research about the quality of the performance of fiscal forecasters and the best practices to provide excellent forecasts. In the literature, fiscal forecasting is considered more an art than a science (Pedregal et al., 2014; Leal et al., 2008), and the importance of statistical rigor is less than the necessity to provide tools to understand the budgetary trends and so, direct the policy maker's decisions. In this sense, the role of the forecasts is at the centre of an intense debate.

A first issue identified by the literature is related to the performance of fiscal forecasters. In particular, the forecasts have been studied in terms of quality and efficiency[1]. Many studies claim that these results contain politically motivated

[1] Provided by public organizations such as National Ministries of Economy and Finance, Central Banks, Statistical National and international Institutes as the International Monetary

biases in projections rather than realistic economic forecasts, offering some good analysis, but also big misses -- especially when big breaks happen. So a huge part of the literature suggests that public finance projections in Europe should be produced by independent agencies to avoid politically motivated and biased forecasts (Perez, 2007, Pedregal et. al., 2014; Abreu, 2011). These studies propose to introduce independent forecasts in the public fiscal area (Debrun et al., 2009, Leeper, 2009, Wypslosz, 2008, Jonung and Larch, 2006a,b, European Commission, 2006). Nevertheless, other studies demonstrate that the bias of independent agencies' fiscal forecasts are generally different from those of the Government's fiscal projections. Indeed, there are some problems of correlations between agencies regarding the forecasts published by private agencies that produce contagion effects between the predictions. Other studies claim that the forecasts produced within Government agencies are more accurate than those produced outside because the former have early access to sensitive data and legislative changes that affect both the financial and economic forecasts and that are not available at the same time to the outside forecasters (Klay and Grizzle, 1992). In line with this discussion, it is important to examine the issue in depth, analysing the efficiency and testing the quality of the fiscal forecasts produced by different institutions.

A second relevant issue regarding fiscal forecasts essentially concerns the search for more accurate methods to identify the best practices in forecasting. Although there are no definitive conclusions on where studies converge, especially when big breaks happen, many researchers are engaged in making continuous choices about forecasting procedures - in particular to make their economic and fiscal models consistent and to approximate as much as possible their predictions to the Government budget. They implement many procedures based on different techniques, including one based on combination. On this topic of investigation, there is much evidence in the literature about combining forecasts that shows how good the results are when using this practice (Clemen, 1989). In general, the benefits of combining forecasts are better than when different methods or theories are used individually (Batchelor and Dua, 1995). These models or methods provided by the different forecasters have the goal of searching for a "true model" with the hope of explaining the best scenario in the future. But as many authors argue (Winkler and Makridakis, 1983) the

Fund (IMF), the Organisation for Economic Co-operation and Development (OECD) or European Commission (EC), (Artis and Marcellino 1998, 2001; Keerman, 1999; Timmerman, 2006, 207) idea of combining is not based on finding a true model for forecasting purposes. Indeed, the world may change at any moment and economists have to consider in their models input that has not been examined previously. This seems to suggest that the models have to be adaptative, but also that they represent the best state of knowledge rather than any absolute truth. So the approach that seems to be more realistic and consistent with the existing literature considers the forecasts provided by different forecasters in conjunction with information about events and variables within a certain period of time. Using this approach, the forecasters are not in competition with each other, but rather the information they provide is used to complement the information provided by other forecasters to form a more complete picture. This is the reason to combine forecasts: to have an aggregation of information that reduces uncertainty and errors and increases accuracy.

This idea has inspired, during the last forty years, many authors in the field to consider the value of combining fiscal forecasts. In their seminal contribution, Bates and Granger (1969) discovered that the simple average of multiple forecasts outperforms that of the forecasts taken individually. The idea was to use the relative combination of variances and covariances to construct a weighted average of the forecasts minimizing the mean square error of the combined forecast. This idea was extended to a huge literature. Clemen (1989) provides a very deep review of the methods used to combine and confirm these results. At the same time, Clements and Winkler (1986) give the idea of combination in their philosophical approach. More recent evidence from Stock and Watson (1999(a), 1999(b), 2002, 2003 and 2004) finds that combination forecasts provide a useful way to incorporate information from a large number of potentially relevant predictors. In the last year, other economists have given their attention to combining fiscal variables: for example, Faust and Wright (2008) on the exchange rates; Andreou et al. (2011) on output growth; and, Ozkan (2011) on deficit/GDP in the USA. To test possible fluctuations in the relative predictive abilities of forecasting models and combination models during the forecast evaluation period, new tests have also been recently proposed; for instance, Giacomini and Rossi, 2010 propose a test that focuses on the entire time path of the model's relative performance.

In line with the previous discussion, there are other lines of research that have been effectively developed in recent years with the aim of improving public finance forecasts. The increasing need to provide financial and economic variables at high frequencies has driven research into the new concept of "nowcasting". In practice, nowcasting is "forecasting" the current or recent aggregate state of an economy using information from data that are related to the target variables but collected at a higher frequency, typically monthly or even real-time (Banbura et al, 2010, 2012; Angelini et.al, 2010). It is important that the policy makers have updated forecasts to make decisions whenever it is necessary and in part these recent studies have this aim. Even if, currently, decisions are based on forecasts and although official and private institutions are engaged in producing scenarios, as much as possible, that are consistent with economic reality through accurate forecasts and adequate fiscal policies, there are different weaknesses that currently affect their capacity to provide a proper analysis of the future evolution of fiscal deficits:

First, it has become increasingly difficult for the analysts to follow and interpret the continuous and inconstant flow of monthly fiscal data that is currently published by official statistical agencies (Perez, 2007). In fact, the publication of the monthly data of the different macroeconomic variables is not contemporary. This means that it is difficulty to set models that consider real-time variables for which data are published in a non-homogeneous way.

Second, the proposed models are full of information and variables that often make them difficult to read and interpret. Often a parsimonious model with a low cost of implementation can be an advantage in terms of understanding the economic reality and its dynamics. A model with fewer variables may be less affected by problems associated with the publication dates of variables and thus more efficient.

Third, the timing of the availability forecasts is relevant (Keerman, 1999). If governments produce forecasts once a year and international agencies such as the EC, the OECD, and the IMF do so twice a year (usually in Spring and Autumn), the vision of the real performance of the economy and its public finance balances cannot be updated. One should have access to the accounts continuously in real time in order to understand how to better adjust fiscal policies and laws to public finance, or vice-versa. This lack of timing availability during the year is typically covered by private analysts who make "adjustments" to their models on a monthly basis. The Consensus Economic Forecast (CEF) collects these data, which, however, are discontinuous for all agencies.

In fact, these data and forecasts are not published for all months, and thus cannot always be taken into account by policy makers.

I.2. Objectives and structure of the thesis

With all this in mind, the main purpose of my doctoral thesis is to provide new insights regarding the analysis and the improvement of the fiscal forecast's performances in Italy, taking into account the three weaknesses identified at the end of the previous paragraph. Italy´s increasingly high debt levels and unbalanced economic deficit record provide a macroeconomic scenario that needs relevant ex ante policy decisions at both the country and the EU level. The attention of policy makers to Italy is explained by the history of public debt. Italian debt has been over 100% of GDP for most of the past two decades. During the years 1994-1996, the debt went over 106% of GDP, then decreased from 1998-2000 and then increased again in 2008 to 127% of GDP in the last few years against the EU average of 68%. Another variable to take into account, which is related to the yearly fiscal performance, is the public deficit. It was 2.6% on GDP in 2014. Regarding the evolution of deficit, from 1992 to 1996 this variable was very high at 10.3% against the EU average of 5.9%, but it decreased by the year 2000 with the introduction of the Euro system. After 2001 it began to increase again until 2007, when it reached 5.4%. It stood at 3.9% in 2010, and has continued decreasing until today[2]. This volatility entails an effort to stabilize the trend of these variables. To keep public finances on the Maastricht targets and optimize the emission of public debt stocks, it is essential for policy makers to have a forecasting tool available. The expected path of the related annual variables, such as the deficit, provides the future trend for budget public finance. The emphasis on the deficit is justified by the fact that this is the key variable describing the state of the "health" of public finance. Monitoring this variable could promptly correct unfavourable trends and possible crisis. Budget forecasts in this way became a crucial part of a democratically controlled policy process. They are now a key input in informed budget drafting and decision-making, and a tool to manage expectations of fiscal responsibility in financial markets and the public at large.

To examine the role of budget forecasts in depth, the thesis consists of three self-contained essays on Italian fiscal forecasting over the last two decades. The aim is to develop different tools for improving forecasts, in order to support the Government in projecting budget plans in general, and the public deficit in particular.

---

[2] Source OECD.

In this way, I propose innovative approaches to fiscal forecasting through performance checks on a large set of forecasters, the combination of many forecasters to improve performance, and the application of nowcasting models. In view of the growing sensitivity of policy makers and researchers to budget forecasts, this thesis finds its reason on a review of the main studies in accuracy of fiscal forecasts, search tools through which improve current practices and tries to implement techniques to provide monthly forecasts that are useful for economic policy making.

Chapter II, "Accuracy of fiscal forecasts in Italy" is focused on one of the most important aspects of the new Treaty: it requires that the decisions and recommendations taken by the European Council no longer be based on outcomes but on forecasts. For countries that have a very high level of debt and large deficits, like Italy, this new procedure is supposed to be particularly helpful. However, implementing this new framework may be a daunting task. The production of fiscal forecasts has been limited to government agencies and international institutions, and private forecasters have neglected budget forecasts altogether as they have not been considered to be key macroeconomic variables. Evaluation of fiscal forecasts has shown that most methods have poor reliability.

In this chapter, I evaluate whether fiscal forecasts for Italy are accurate and econometrically efficient, and if so, whether these forecasts be used by the EC to make recommendations to member states. Similar ex-post accuracy tests of budget forecasts have been carried out for different countries (Von Hagen, 2010, Jonung and Larch, 2006; Pina and Vanes, 2011). These tests use a limited number of forecasts from government agencies. In contrast, I focus on a large number of deficit forecasts for Italy that come from a variety of sources, including both public and private agencies as well as Italian and international institutions. I analyse the extent of the discrepancies between the yearly released deficit on GDP and its forecast in Italy from January 1992 to December 2011. The fiscal forecast records come from international (IMF, OECD, and EC), private (bank, companies, research institutes etc.) and public national organizations such as the Italian Ministry of Economy and Finance (MEF). I conduct two types of analysis: first, I analyse the forecasts of each individual agency from a quatitative perspective, and then I conduct qualitative general investigation of the same forecasts of each agencies. In the first type of analysis, I carry out different accuracy tests to detect which organization is the best forecaster and in what part of the year better results are published. I also compare forecasters' performance against a naïve benchmark model, which provides a minimum level of accuracy as in Marcellino (1998, 2001), Keereman (1999) or Gordo and Martins (2007). In the second type of analysis, I consider the quality of the forecasts and I test weak efficiency, unbiasedness, and serial correlations. Following Artis and Marcellino (1998, 2001) and Keereman (1999), I deepen the analysis for Italy, using a new database and considering different kinds of forecasters. I conclude that all fiscal forecasters for Italy provide unbiased and efficient forecasts with very few exceptions. In general, forecast errors do not persist in a regular way, and predictions are efficiently made in the sense that the information included in past mistakes is taken into account. In particular, deficit forecasts tend to overpredict the real deficit value. The most relevant result of this analysis is that private forecasters are frequently more accurate than national and international ones. The EC is often the "best" forecaster amongst different international forecasters (Keereman, 1999, Artis and Marcellino, 1998, 2001).

In Chapter III, "Combine to compete: improving fiscal forecast accuracy over time", I argue that budget forecasts are increasingly becoming a tool of fiscal management as the Financial Crisis led to a fiscal meltdown in developed economies. I argue in this Chapter that I can take advantage of the information contained in all individual budget forecasts analysed in the previous chapter to improve their accuracy. I do this by projecting combined forecasts through pooling the judgment and expertise of the forecasters. In the forecasting literature, it is an established finding that combining improves upon the forecast of any single model (Winkler and Makridakis, 1983, Clemen, 1989, Clemen and Winkler 1986, Hendry and Clements, 1998, 2003, Timmerman 2006, Hendry and Hubrich, 2011). Following this idea of improving the forecasting accuracy, I apply a variety of combination techniques, both simple and advanced, which account for past forecasting performance, to compute a combined forecast. I look at a dataset of nine monthly expert forecasts from private agencies and semi-annual projections by public institutions for Italy over the period from 1992 to 2012, which was analysed in the previous Chapter. My main finding is that different combinations of budget forecasts often result in more accurate forecasts than individual models. This is particularly the case for a weighted forecast combination and Rbest that value the forecasts that have been more accurate in recent periods. Standard tests of forecasting accuracy show that even one year ahead, some of the pooled forecasts significantly outperform a naïve model, which is a substantial improvement over expert forecasts (Artis and Marcellino, 2001) or over a variety of forecasting methods (Favero and Marcellino, 2005). Substantial improvements on fiscal projections are possible by using a set of budget forecasts and checking their performance. Although the constant follow-up of forecast performance helps in improving accuracy, structural changes make predictive accuracy challenging over time. I use recently developed tests to check forecasting accuracy over time (Giacomini and Rossi, 2010), and find that the weighted forecast combination outperforms other predictors over all years. Its improvement in accuracy is statistically significant when compared to a naïve model.

Chapter IV, "Nowcasting public finance in Italy," moves from the idea of forecast and combination of annual data to the most recent idea of nowcasting fiscal variables. The reason is to give policy makers the capacity for dynamic monitoring of the public budget's cash flow. This monthly analysis exploits the information at higher frequencies before the official figure becomes available. The approach that I use consists of using different nowcasting techniques that are well known in the literature. In particular, I propose a set of models that are parsimonious and suitable for real-time monitoring of the fiscal deficit. The purpose of this work is to make available forecasts of fiscal deficits in those months in which official forecasts are not published. This study can be an excellent tool for policy makers to adapt policies to the real economic situation of the country and avoid spirals of debt and crisis. In this Chapter, I conclude that the linear regression models outperform the other techniques used. The introduction of public finance and economic confidence variables and Google trends results in performance gains when compared with the time series and autoregressive models.

Finally, Chapter V, "Conclusions" provides a general outlook of the previous three chapters, describes the main conclusions of the entire thesis, and offers related policy implications and a discussion of further research.

I.3. References

Abreu, L. (2011). *International Organizations's vs. private analysts' forecasts: an evaluation* (Banco de Portugal working paper No. 20).

Andreou, E., Ghysels, E., & Kourtellos, A. (2011). Forecasting with mixed-frequency data. In M. Clements & D. Hendry (Eds.), *Oxford handbook of economic forecasting*. Oxford University Press.

Angelini, E., Banbura, M., & Rünstler, G. (2010). Estimating and forecasting the euro area monthly national accounts from a dynamic factor model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, *2010*(1), 1-22.

Artis, M., & Marcellino, M. (1998). *Fiscal solvency and fiscal forecasting in Europe* (CEPR Discussion Papers No. 1836)

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal*, *4*, 20-36.

Banbura, M., Giannone, D., & Reichlin, L. (2010). *Nowcasting* (CEPR Discussion Papers No. 7883).

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2012). Now-Casting and the Real-Time Data Flow (Universitée Libre de Bruxelles Working Paper No. 26).

Batchelor, R., & Dua, P. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science, 41*(1), 68-75.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20,* 451-468.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*, 559-583.

Clemen, R.T., & Winkler, R.L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics, 4*, 39-46.

Clements, M.P., & Hendry, D.F. (1998). *Forecasting economic time series.* Cambridge: Cambridge University Press.

Clements, M. P., & Hendry, D. F. (2003). Economic forecasting: some lessons from recent research. *Economic Modelling, 20*(2), 301-329.

D'Agostino, A., McQuinn, K., & O'Brien, D. (2011). *Nowcasting Irish GDP* (University of Munich MPRA Paper No. 32941).

Debrun, X., Hauner, D. & Kumar, M.S. (2009). Independent fiscal agencies. *Journal of Economic Surveys* 23, 44-81.

Faust, J., & Wright, J.H. (2008). *Efficient prediction of excess returns* (National Bureau of Economic Research NBER Working Papers No. 14169).

Favero, C., & Marcellino, M., (2005). *Modelling and forecasting fiscal variables for the Euro Area* (Bocconi University Working Papers No. 298).

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25,* 595–620.

Gordo Mora, L., & Nogueira Martins, J. (2007*). How reliable are the statistics for the Stability and Growth Pact?* (Directorate General of Economic and Financial Affairs, European Commission, European Economy Economic Papers No. 273).

Hameed, F. (2005). *Fiscal transparency and economic outcomes* (IMF Working Paper No. 225).

Hendry, D. F., & Clements, M. P. (2003). Economic forecasting: some lessons from recent research. *Economic Modelling, 20*(2), 301-329.

Hendry, D. F., & Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics, 29,* 216–227.

Jonung, L., & Larch, M. (2006a). Fiscal policy in the EU: are official output forecasts biased? *Economic Policy*, July, 491-534.

Jonung, L., & Larch, M. (2006b). Improving fiscal policy in the EU: the case for independent forecasts. *Economic Policy, CEPR; CES; MSH, 21*(47), 491-534.

Keereman, F. (1999). *The track record of the Commission forecasts* (European Commission, Directorate-General for Economic and Financial Affairs Economic Paper No. 137).

Klay, W. E., & Grizzle, G. (1992). Forecasting state revenues: Expanding the dimensions of budgetary forecasting research. *Public Budgeting and Financial Management, 4*(2), 381-405.

Leal, T., Pérez, J. J., Tujula, M., & Vidal, J. P. (2008). Fiscal forecasting: Lessons from the literature and challenges. *Fiscal Studies, 29*, 347–386.

Leeper, E. M. (2009). *Anchoring Fiscal Expectations* (NBER Working Paper No. 15269).

Ozkan, N. (2011). *A first step to develop a forecasting model for U.S. Government budget deficit: A regression-based method with quarterly economic indicators*. University of North Carolina at Chapel Hill, Department of Economics.

Pedregal, D. J., Pérez, J. J., & Sánchez-Fuentes, A. J. (2014). A toolkit to strengthen government budget surveillance (Banco de España working paper No. 1416).

Pérez, J. J. (2007). Leading indicators for euro area government deficits. *International Journal of Forecasting 23 (2)*, 259-275.

Pina, A., & Venes N. (2011). The political economy of EDP fiscal forecasts: an empirical assessment. *European Journal of Political Economy, 27*, 534-546.

Stock, J., & Watson, M. (1999a). Forecasting inflation. *Journal of Monetary Economics, 44*, 293–335.

Stock, J., & Watson, M. (1999b). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In Engle R, & White, H. (Eds.), *Cointegration, Causality and Forecasting, A Festschrift for Clive W.J. Granger* (pp. 1–44). Oxford: Oxford University Press.

Stock, J., & Watson, M. (2002). *Combination forecasts of output growth and the 2001 Recession*, Mimeo, Harvard University and Princeton University.

Stock, J., & Watson, M. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Perspectives, 41*, 788–829.

Stock, J., Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 405-430.

Timmermann, A. (2006). Forecast Combinations. In Elliot, G., Granger, C. & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 136-196). North-Holland, 1,

Timmermann, A. (2007). An evaluation of the World Economic Outlook forecasts. *IMF Staff Papers, 54 (1),* 1–33.

Von Hagen, J. (2010). Sticking to fiscal plans: the role of institutions. *Public Choice, 144,* 487-503.

Winkler, R.L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, 146*, 150-157.

Wyplosz, C. (2008). Fiscal policy councils: Unlovable or just unloved. *Swedish Economic Policy Review*, *15*, 173-192.

Chapter II.

Accuracy of fiscal forecasts in Italy [3]

II.1. Describing fiscal forecasts

Budget forecasting may look like the exclusive task of Ministries and international institutions. Yet many other expert forecasters, like commercial or investment banks, industry, semi-governmental agencies, and university departments have produced budget forecasts too. In recent years, some datasets have become available that include deficit forecasts from a larger set of expert forecasters over a continuous period of time. One of those datasets comes from Consensus Economics Forecasts (CEF). This company conducts a survey in several OECD countries among professional economists working for commercial or investment banks, industry, government based agencies, and university departments. Most of the surveyed experts are at domestic institutions that provide forecasts for a single country only; a few work for international financial institutions or research institutes that provide forecasts for several countries simultaneously.

The CEF survey has gradually expanded its scope and coverage, and provides us with a large panel of private forecasters. The monthly survey on Italy covers 42 forecasters from 1992 to 2012. However, despite the gradual expansion of the dataset, fiscal forecasts have not always received the same attention by forecasters over time. Some forecasters stopped producing projections for the

[3] I presented this chapter to the AQR Group at the Universitat de Barcelona and to CER - Centro Europeo di Ricerche - in Rome, (Italy), which is one of the private forecasters in the analysis and to the Ministry of Economics and Finance of Italy Tresury's Department, which is one of the public forecasters in the analysis in July 2013. I presented at the PhD day at the Universitat de Barcelona and at the 14th IWH-CIREQ Macroeconometric Workshop in Halle (Germany) in December 2013. I presented at Xrepp PhD day at the Universitat de Barcelona in March 2014 and in the Workshop of Time Series Econometrics in Zaragoza (Spain) in April 2014. The paper was also accepted at the Prague Macroeconomics and Finance Conference (Czech Republic), the International Conference on Applied Business and Economics (ICABE) New York (USA), International Academic Conference in Istanbul, (Turkey), XVII Encuentro de Economía Aplicada in Gran Canaria (Spain), Time Series-ITISE 2014 University of Granada (Spain).

budget balance, while others that were initially included left the sample owing to closure, mergers, or other reasons. Moreover, new forecasters joined the CEF survey only at a later stage.

The survey makes enquiries of respondents the first week of each month about current and year ahead forecasts for a number of macroeconomic variables. Forecasts of longer horizons are not included, but may not be interesting as the literature shows that budget forecasts are surrounded by substantial uncertainty and large biases (Favero and Marcellino, 2005). The Consensus forecasts are published early in the second week of the same month.

As mentioned in Chapter I, the reforms of the European fiscal framework have made fiscal forecasts a centrepiece of macroeconomic policy decisions, particularly in highly indebted European Union countries such as Italy. In light of this, I analyse the performance of different forecasting institutions (national/international and public/private) in predicting the Italian fiscal deficit during the last two decades. With this aim, a new data set on fiscal forecast data for Italy has been compiled and standard forecasting competition methods are applied. My sample is therefore a subset of the entire group of expert forecasters included in the CEF survey. I do not consider those forecasters that have participated just a few times in the survey. In particular, any forecaster participating fewer than 12 consecutive months in the CEF survey is excluded. This reduces the panel to a selection of five forecasters among Italian banks and research institutes. To preserve the confidentiality of the respondents, I call these forecasters N1 to N5. I also analyse the data provided by national-level public agencies, such as the Ministry of Economy and Finance (MEF), and international organizations, such as the Organisation for Economic Co-operation and Development (OECD), the European Union (EU), and the International Monetary Fund (IMF). I do this by testing the quality of these forecasts with the RMSE, the MSE, the MAE, and the Theil Index with the naïve model and the Diebold and Mariano test. Then, I test the efficiency and unbiased with the Wald test, and serial correlation with the Bosch and Lagrange multiplier test, considering the forecasts individually by month. I conclude that the accuracy of the forecasts mainly depends on the month in which the forecasts are realized and on the nature of the institution making the economic forecasts.

The rest of the Chapter is organized as follows: Section II.2 provides information on the database used for the study, the variables considered, and their different calendar availability. In Section II.3, I analyse individual forecasts using accuracy tests, and compare them with current year and year ahead forecasts from the naïve model. In Section II.4, I econometrically test the unbiasedness, efficiency, and serial correlation of the forecasts taken as a whole and, finally, Section II.5 summarises the main conclusions of the Chapter.

II.2. Description of the dataset

In this section, I describe the series representing actual data and forecasts of the deficit to GDP ratio, from 1992 to 2012. The actual series is the registered deficit to GDP ratio (d_t) from the OECD's database (OECD iLibrary- Economics: Key tables from OECD). Regarding forecasts, I construct deficit forecasts for both the current year ($d_{f,t}$) and the year ahead ($d_{f,t+1}$) over the sample period from 1992 to 2012. The forecasts require some transformation before they can be used in the empirical analysis. CEF asks respondents for a forecast of the overall balance in nominal terms.[4] In order to transform this forecast into one of the budget balance as a ratio to GDP, I divide the forecast of the nominal balance (surplus) for year $t+1$ in a certain month m by the GDP forecast for the same year. As the CEF only provides forecasts of GDP *growth rates*, I compute the year ahead nominal GDP forecast by applying the CEF growth rate to the latest available estimate for the same year GDP.

I select 5 forecasters out of the 24 private forecasters[5] present in the CEF database because they meet the data requirements necessary, in terms of having sufficient data to conduct my investigation[6]. In addition to the private forecasts,

---

[4] For Italy, specialists forecast the general budget balance for the calendar (end of the) year.

5 The private forecasters include Banca Commerciale, Banca di Roma, Banca IMI, Banca Intesa, Banca Nazionale del Lavoro, Bank of America, Caboto, Capitalia, Cariplo Spa, Centro Europa Ricerche, Chase Manhattan – Milan, Citigroup, Cofiri SIM, Confindustria, Credito Italiano, Deutsche Bank, Euromobiliare, ENI, Econ Intelligence Unit, Fiat SpA, Goldman Sachs, FAZ Institut, ING Financial Markets, HSBC, IHS Global Insight, IRS, ISAE, Intesa Sanpaolo, ISCO, Istituto Bancario Italiano, JP Morgan, Prometeia, Morgan Stanley, RASFIN, Salomon SB Citibank, Studi Finanziari, Schroder SSB Citibank, UBS, UniCredit, UniCredit Banca Mobiliare.

[6] I consider only those forecasters that produce, in the sample, more than 10 consecutive observations.

I also consider public deficit forecasts for the current year ($d_{f,t}$) and the year ahead ($d_{f,t+1}$). These forecasts come from four institutions: the OECD, the IMF, the EC, and the Italian Ministry of Economy and Finance (MEF). The international institutions do not produce forecasts on a monthly basis. Instead, generally speaking, they produce their projections twice a year (in Spring and Autumn), at different moments. The OECD publishes its forecasts in June and December in the Economic Outlook; the IMF forecasts are published in the World Economic Outlook; and forecasts by the EC are released in May and October. The publication of forecasts by the Italian Ministry of Economy and Finance is part of the "Economic and Financial Planning Document (DPEF)" from 1992 to 1997, and the "Forecast and Planning Report (RPP)" from 1998 to 2012 that are used by the Italian Government when submitting the budget to Parliament. These forecasts are produced in June, July, and October.

Table II.1. Timing of release of deficit forecast

Month	**Current year forecast d_{ft}**	**Year ahead forecast d_{ft+1}**
May	EC IMF Private forecasters (CEF, N_1 to N_5)	EC IMF Private forecasters (CEF, N_1 to N_5)
June	OECD Private forecasters (CEF, N_1 to N_5)	OECD Private forecasters (CEF, N_1 to N_5)
October	MEF EC IMF Private forecasters (CEF, N_1 to N_5)	MEF EC IMF Private forecasters (CEF, N_1 to N_5)
December	OECD Private forecasters (CEF, N_1 to N_5)	OECD Private forecasters (CEF, N_1 to N_5)

Note: MEF projections are published in July during 1992-1995, June in 1996-1997 and October during 1998-2012.

Table II.1 shows how I match the timing of the four public forecasters with the five CEF forecasts. I can match four months where there is a correspondence between the nine forecasters (May, June, October, and December).

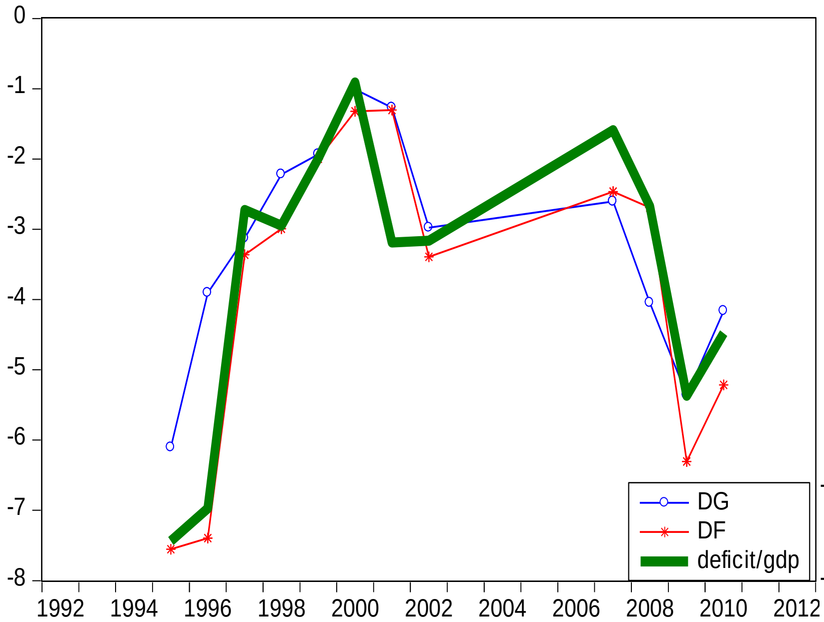
II.3. Analysis of individual forecasts for the current year ($\mathbf{d_{f,t}}$) and year ahead forecast ($\mathbf{d_{f,t+1}}$)

For reasons related to the confidentiality of the respondents, I do not show the real name of the private forecasters; instead I assign a name from N_1 to N_5 ("N" from National). For public agencies, I use MEF, EC, OECD, and IMF. Figure II.1 shows $d_{f,t}$ and $d_{f,t+1}$ for every single forecaster while Figure II.2 shows $d_{f,t}$ and $d_{f,t+1}$ for the months in which they were made available by the forecasters. As can be seen in Figure II.1, current year forecasts are always closer to the realised deficit at year t than when compared to the year ahead forecast with the realised deficit at year t+1. Despite this, $d_{f,t}$and $d_{f,t+1}$ follow the same pattern. This indicates that the additional information gained from the addition of another month is indeed useful in forecasting the deficit ratio, which is as can be expected. Figure II.2 shows that December is the month in which all the forecasters are closer to the actual deficit compared with the other months for both current and year ahead.

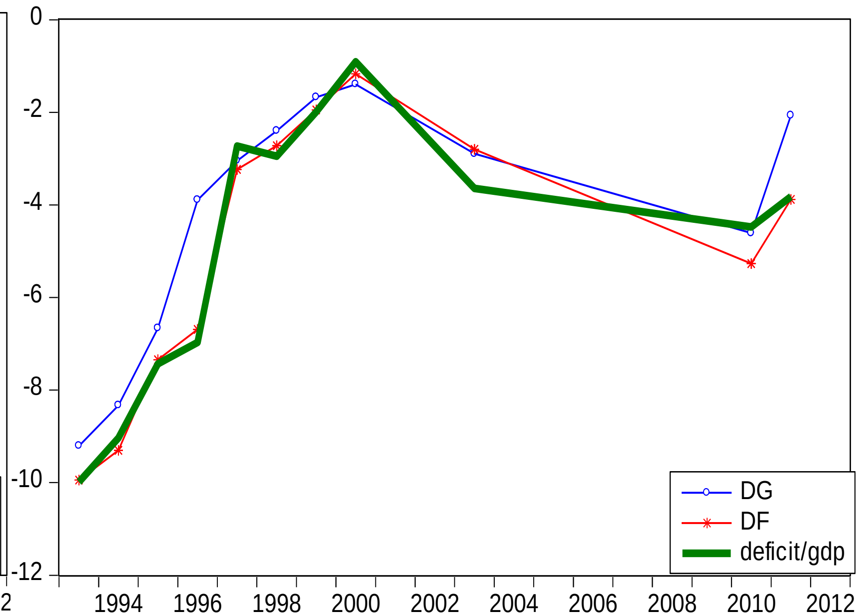
Additional information provided by these figures comes from the fact that when the series $d_{f,t}$ and $d_{f,t+1}$ are above or under the series of real data, this determines if there are overpredictions or underpredictions. So from an economic point of view, overprediction means that the deficit is better than expected, while underprediction means the opposite. Notice that in the official publications a negative value indicates a fiscal deficit while a positive value indicates a surplus. However, throughout the Chapter, because I consider the absolute deficit to GDP values, I will have only positive values. For example, as shown in Figure II.1 and II.2, the comparison between current year and year ahead forecasts indicate that the former are smaller than the latter for all forecasters. In particular, the figures show that both $d_{f,t}$ and $d_{f,t+1}$ have the same pattern in terms of predictions. For example for both $d_{f,t}$ and $d_{f,t+1}$, forecaster N_1, EC and MEF all predicted very close to real data until 2001 where the former began to overpredict and the latter two began underpredicting. N_5, OECD, N_3, and MEF overpredict until 2006-2007; from this point, they start to underpredict while N_4 and N_1 seem to always overpredict.

Figures II.1. Forecast current year, forecast year ahead and deficit/GDP for each forecaster

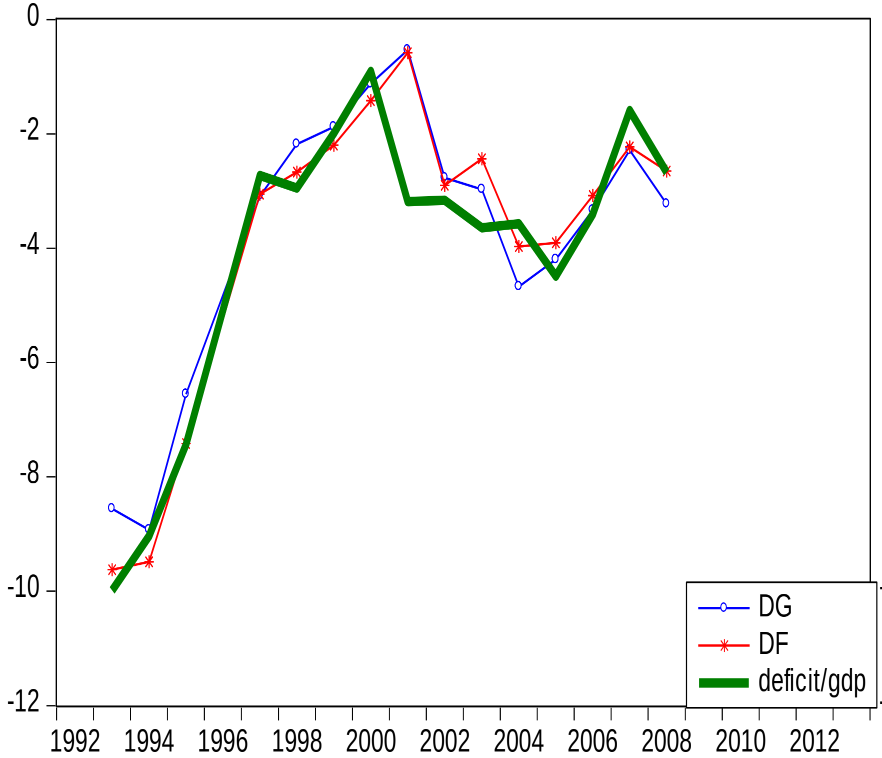
NATIONAL 1



NATIONAL 2

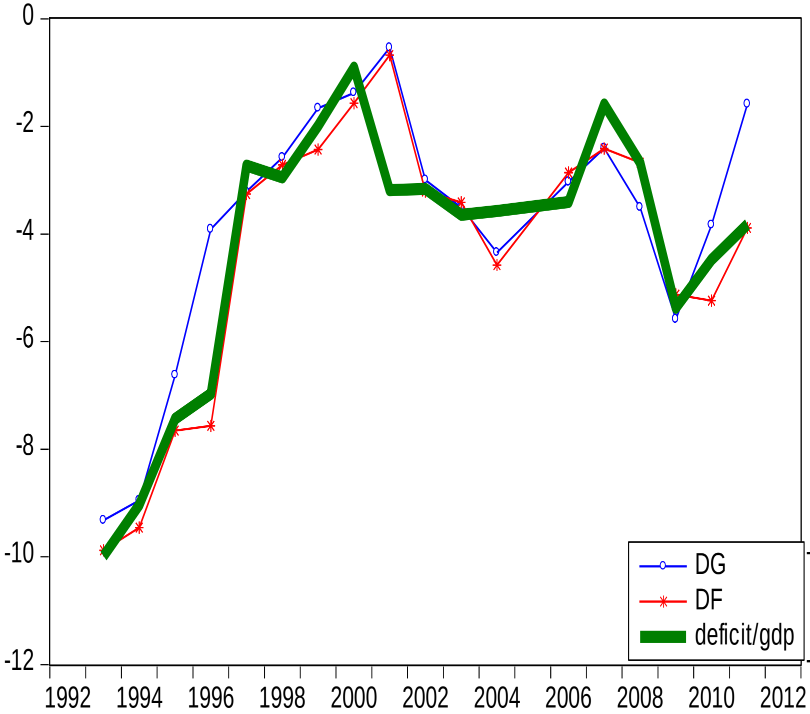


NATIONAL 3

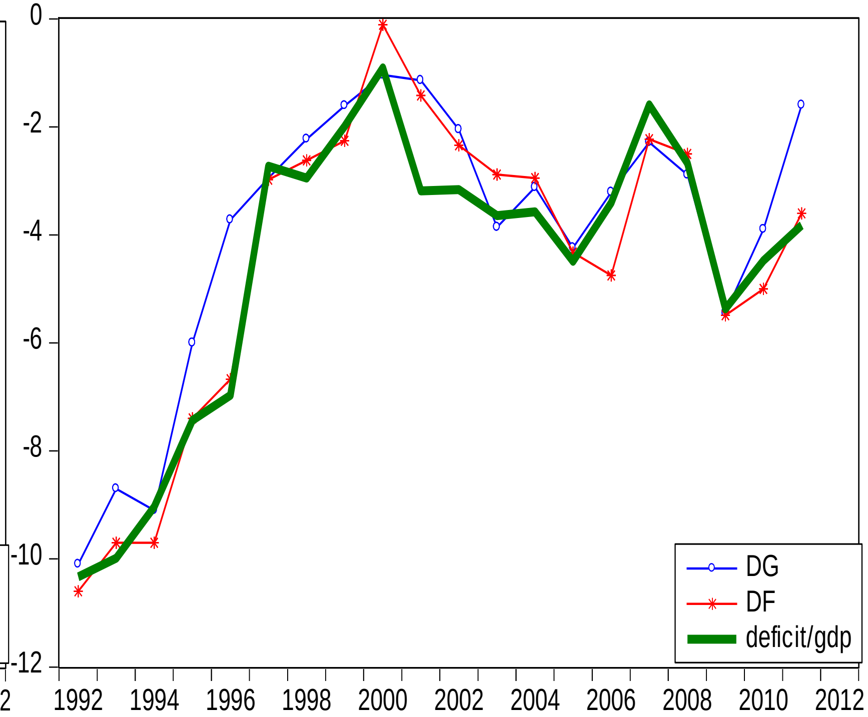


NATIONAL 4

NATIONAL 5



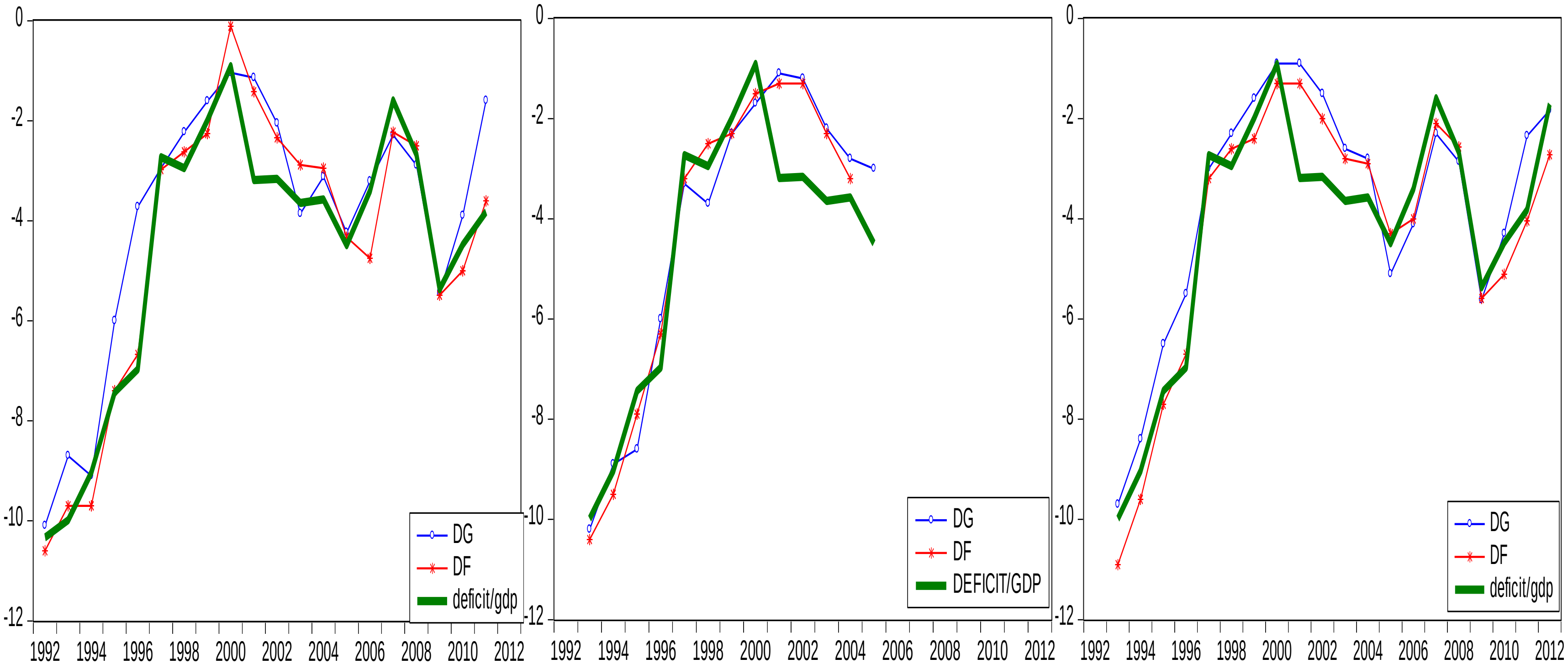
MEF



OECD

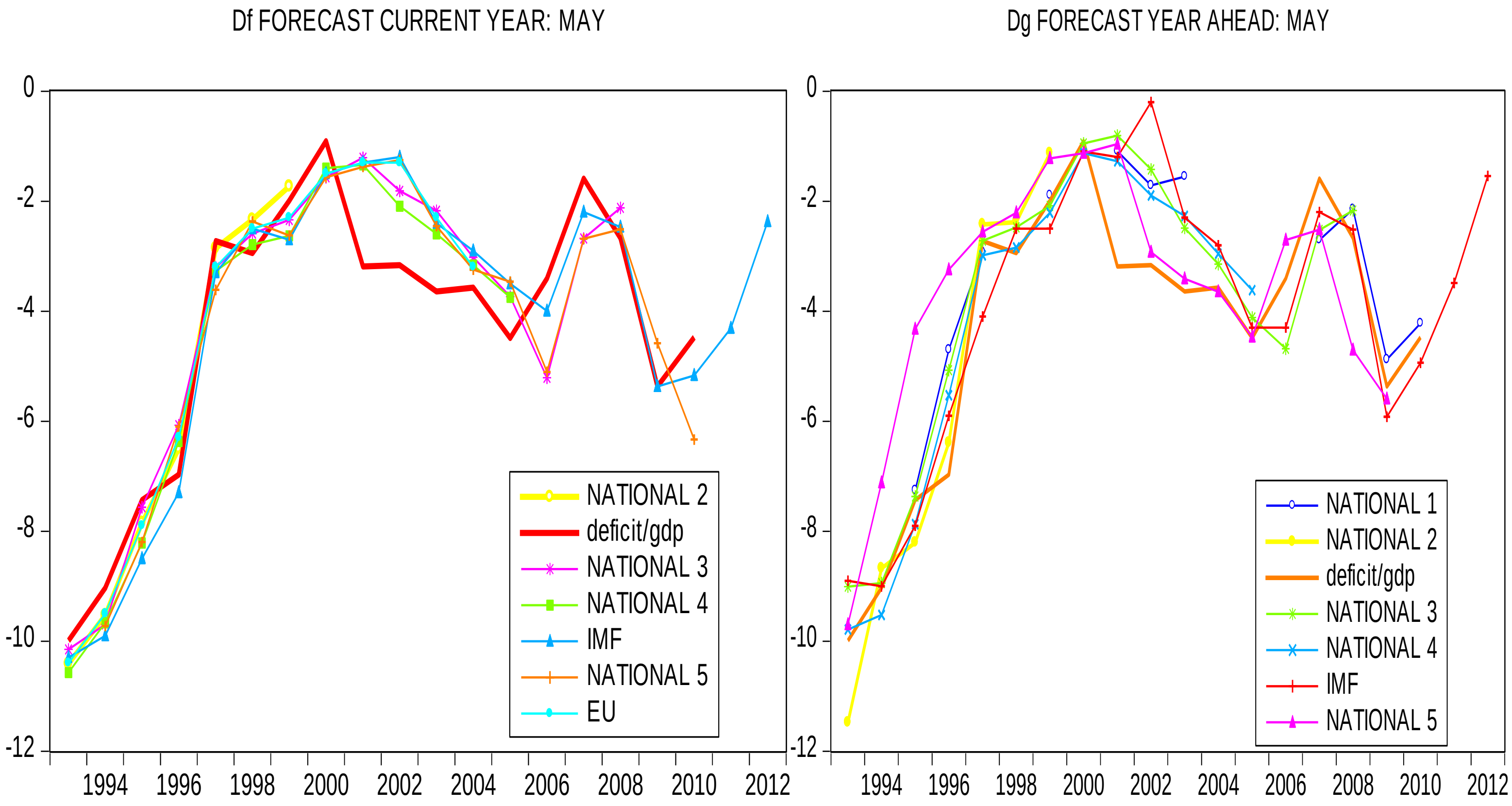
EC

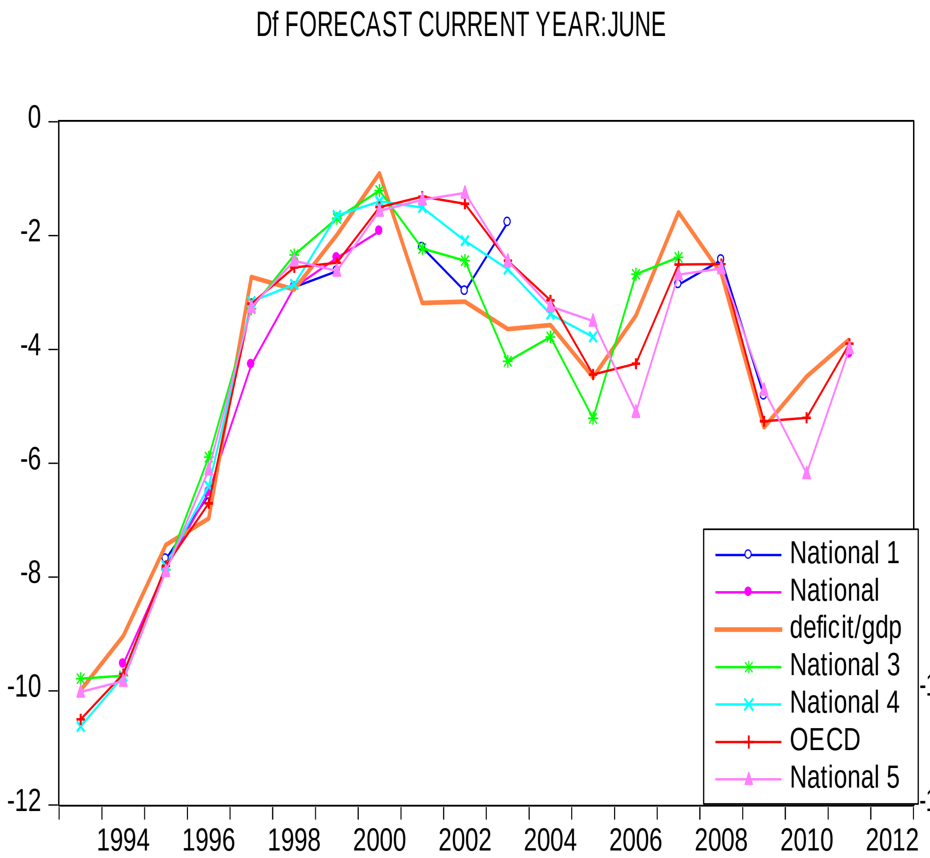
IMF



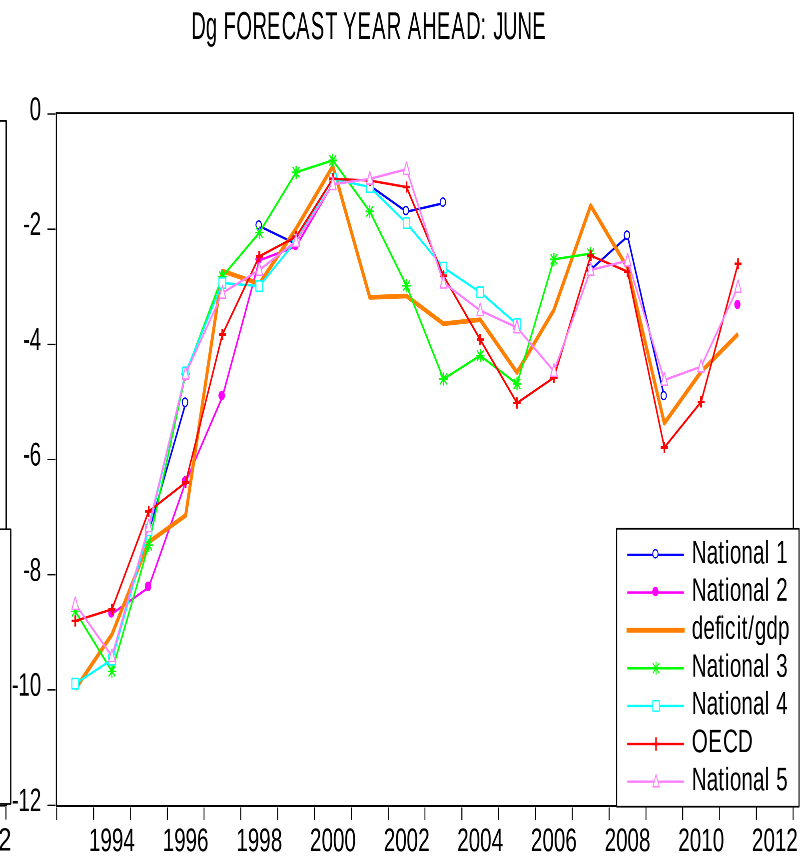
Note: Each forecaster is considered depending on the released data

Figures II.2. Forecast current year, forecast year ahead and deficit/GDP by month

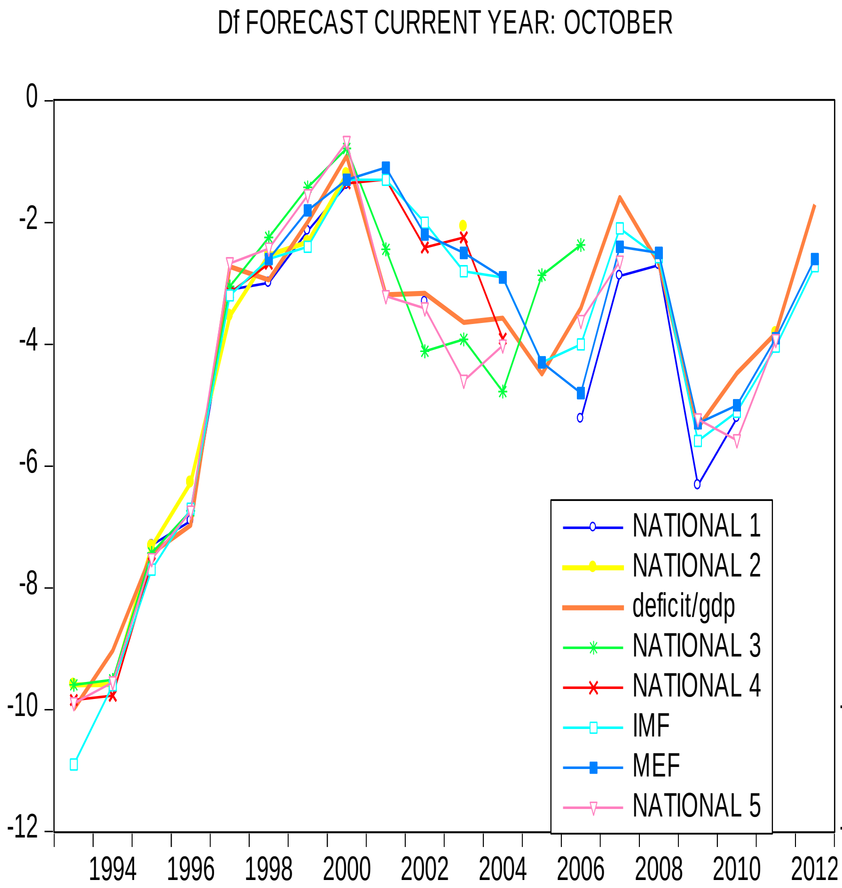


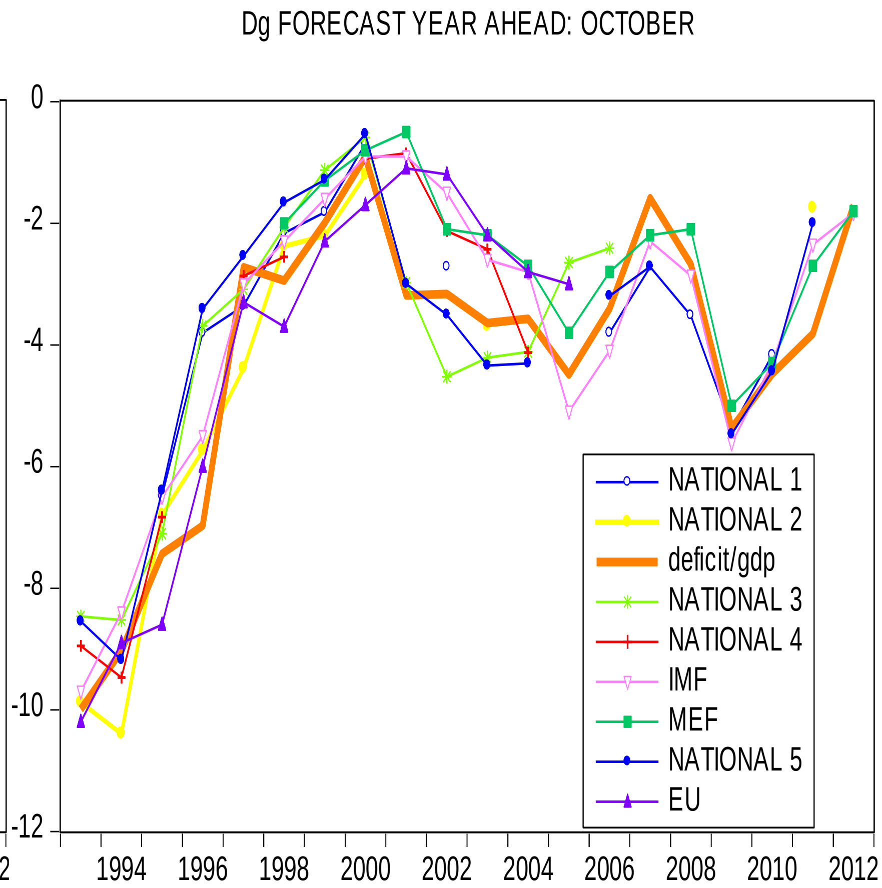


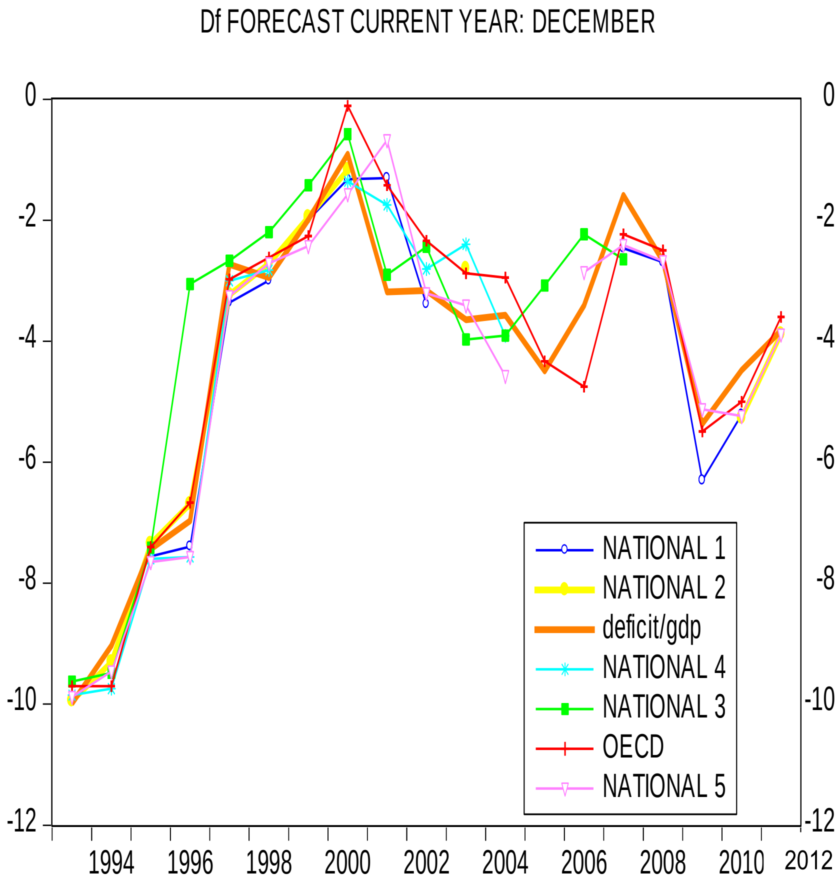
Df FORECAST CURRENT YEAR:JUNE
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
National 1
National
deficit/gdp
National 3
National 4
OECD
National 5

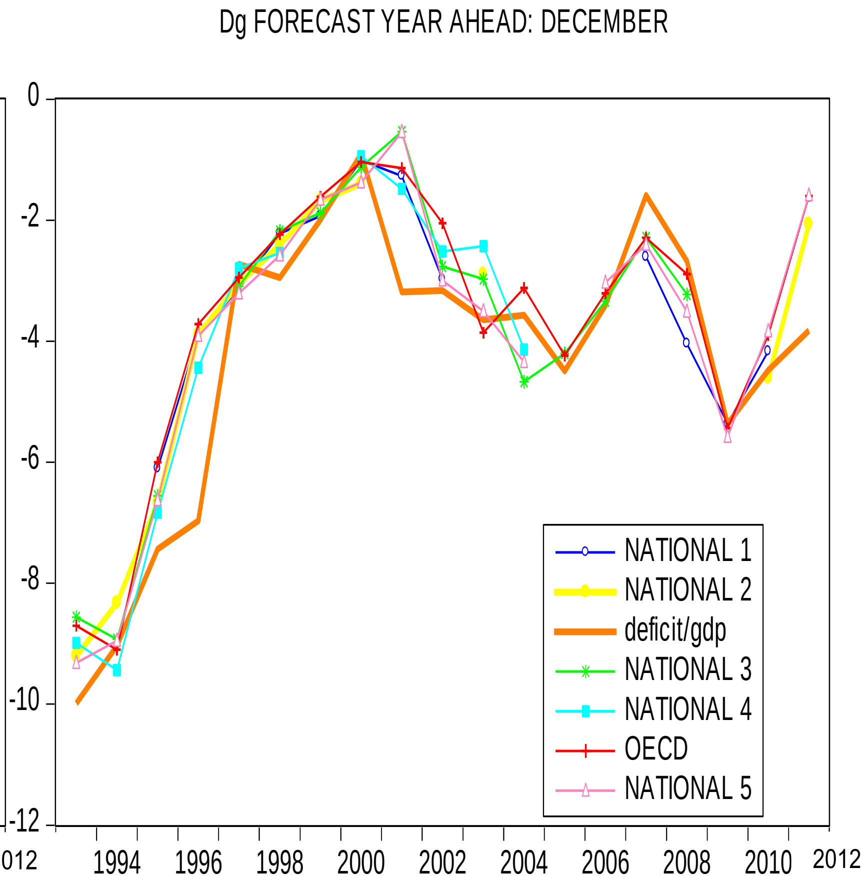


Dg FORECAST YEAR AHEAD: JUNE
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
National 1
National 2
deficit/gdp
National 3
National 4
OECD
National 5


Df FORECAST CURRENT YEAR: OCTOBER
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
NATIONAL 1
NATIONAL 2
deficit/gdp
NATIONAL 3
NATIONAL 4
IMF
MEF
NATIONAL 5


Dg FORECAST YEAR AHEAD: OCTOBER
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
NATIONAL 1
NATIONAL 2
deficit/gdp
NATIONAL 3
NATIONAL 4
IMF
MEF
NATIONAL 5
EU


Df FORECAST CURRENT YEAR: DECEMBER
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
NATIONAL 1
NATIONAL 2
deficit/gdp
NATIONAL 4
NATIONAL 3
OECD
NATIONAL 5


Dg FORECAST YEAR AHEAD: DECEMBER
0
-2
-4
-6
-8
-10
-12
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
NATIONAL 1
NATIONAL 2
deficit/gdp
NATIONAL 3
NATIONAL 4
OECD
NATIONAL 5

II.4. Forecast analysis methodology

II.4.1. Assessing forecast errors

In this section I describe the techniques used to assess forecast quality. There are some forecasters who make more accurate analyses than others. Forecast error is defined as the difference between the actual deficit d_t at time t (and $d_{t+1}$the actual deficit at time t+1) and the forecasted value $d_{f,t}$ at time t (and $d_{f,t+1}$ at time t+1, respectively):

$$e_{f,t} = d_t - d_{f,t} \tag{II.1}$$

$$e_{f,t+1} = d_{t+1} - d_{f,t+1} \tag{II.2}$$

I compute different accuracy statistics based on both forecast errors:

1. ME mean error

$$\text{ME}=\frac{1}{n} \sum_{t=1}^{n} e_t \tag{II.3}$$

Following Keereman (1999), the mean error is equal to the mean forecast minus the realized average. The drawback of this measure is that positive and negative errors can offset each other and thus reduce the size of the error. The mean squared error takes this into account.

2. MSE Mean squared error

$$MSE = 1/n\sum_{t=1}^{n} e_t^2 \tag{II.4}$$

While with the mean error positive and negative deviations of the projection from the actual data can cancel out, this is not the case with the MSE. A MSE of zero means that there is perfect accuracy.

3. MAD Mean absolute deviation

$$\text{MAD}= 1/\text{n} \sum_{t=1}^{n} |e_t| \tag{II.5}$$

The mean absolute deviation is a measure of dispersion. It measures the size of the difference between the values in the projections and the values associated with the actual oucome. Since the absolute value is used, this prevents deviations with opposite signs from cancelling each other out.

4. RMSE root mean squared error of expected budget deficits in terms of GDP:

$$\text{RMSE}=\sqrt{\text{MSE}} \tag{II.6}$$

Large errors are usually considered more harmful than small differences between forecasts and real data. To penalize large mistakes, a root mean squared error (RMSE) can be used. The RMSE is frequently used as a measure of the difference between the predicted values and the values that are actually observed. These individual differences are also called residuals and the RMSE serves to aggregate them into a single measure of predictive power. A large RMSE indicates a lower level of accuracy.

5. Theil's inequality coefficient:

$$\text{T}=\frac{\text{RMSE } e_t}{\text{RMSE naive model}} \tag{II.7}$$

The value of a forecast should not only be appreciated in terms of its own errors, but also compared to the errors of alternative models. The Theil T statistic (Theil, 1971) compares each forecast with a naïve no-change forecast. For example, in the case of five-year averages, this means that the average of the past five years is taken as the benchmark forecast for the outcome in the following five-year period. The Theil coefficient will take the value 1 under the naïve forecasting method. Results of Theil index lower than 1 indicate greater forecasting accuracy by the agency compared with the naïve forecasts, while values greater than 1 indicate the opposite.

6. Diebold-Mariano Test:

$$DM= \frac{\frac{1}{T}\Sigma_1^t\{g(e_{A,t+h})-g(e_{B,t+h})\}}{\hat{\sigma}_{[\![g(e]\!]_{A,t+h})-[\![g(e]\!]_{B,t+h})}} \quad \text{(II.8)}$$

I further apply the Diebold-Mariano (1995) test of predictive accuracy and compare each forecast with a simple naive model. The DM test supposes that a forecaster has an identical loss function g (A,B) so that two different forecasts, A and B, lead to similar losses due to errors. Let $g(e_{A,t})$ and $g(e_{B,t})$ denote the loss from a forecast error evolving from a prediction model A and B, with $\hat{\sigma}$ denoting a consistent estimate of the standard deviation of the difference of losses. The null hypothesis is that $g(e_{A,t+h})=g(e_{B,t+h})$, and DM is simply distributed as N(0,1) under this null (Diebold and Mariano, 1995).

II.4.2 Accuracy of forecast errors

The results for forecasting the accuracy of the error in the year ahead ($e_{f,t+1}$) and in the current year ($e_{f,t}$) are shown in the following figures, in particular, for the ME (Figures II.3. and II.4.), MSE (Figures II.5. and II.6.), RMSE (Figures II.7. and II.8.), MAD (Figures II.9. and II.10.), Theil's index (Figures II.11. and II.12.), and Diebold and Mariano test (Figures II.13. and II.14.).[7] These figures allow us to compare the statistical value of each forecaster with respect to the month the forecast was published.

[7] The detailed results can be found in Annex II, Table A.II.1. for current year and A.II.2. for year ahead.

Figure II.3. Mean of current year forecast

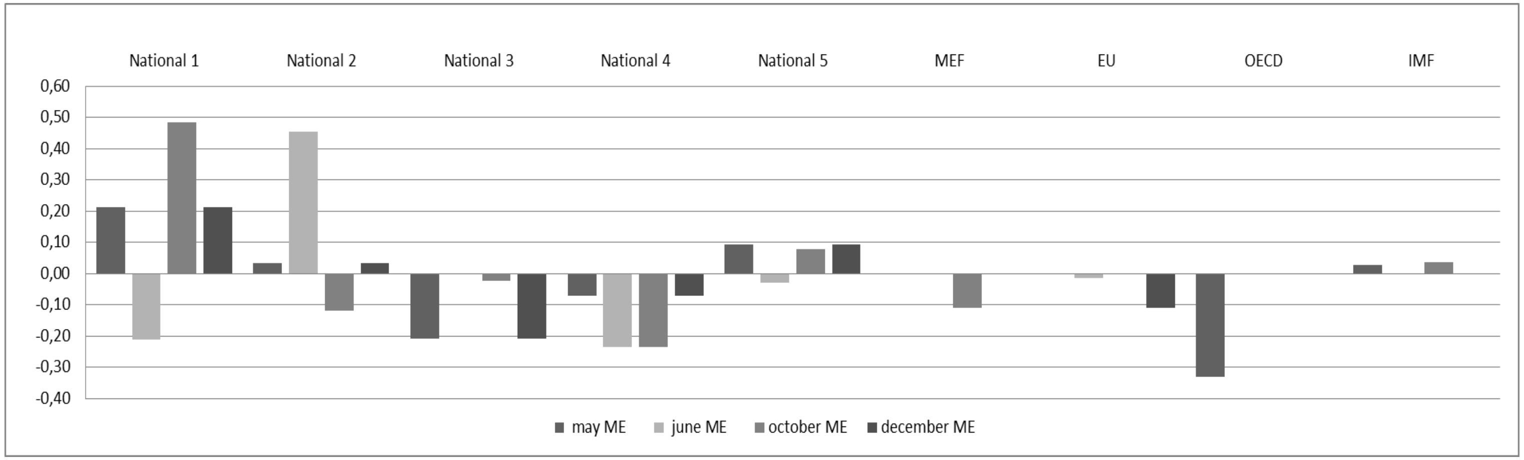


Figure II.4. Mean of year ahead forecast

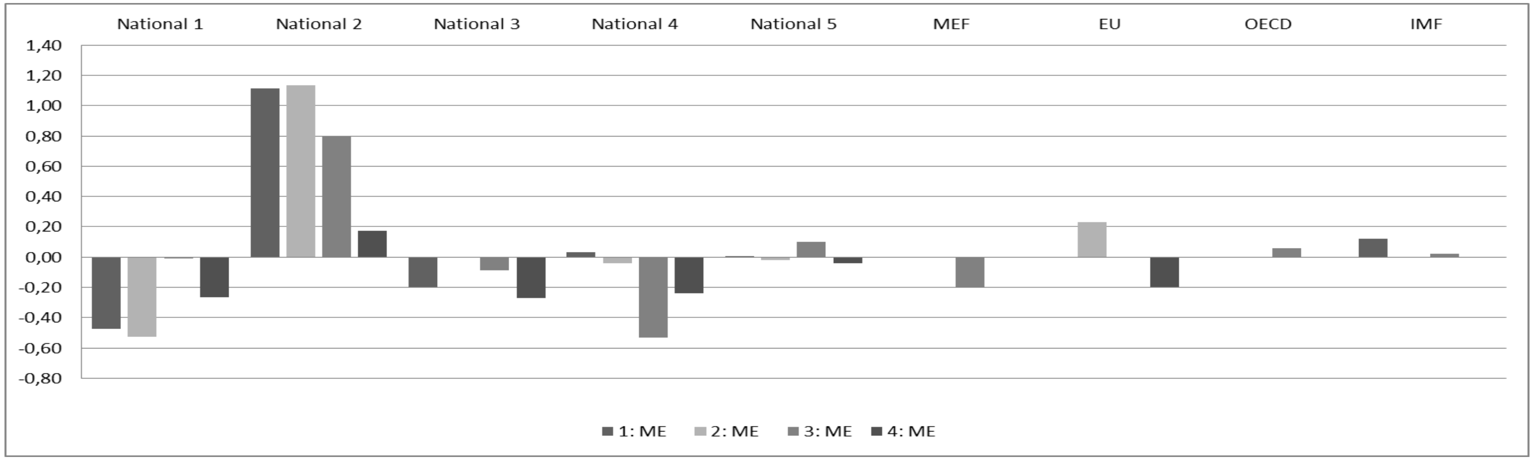


Figure II.5. Mean squared errors of current year forecast

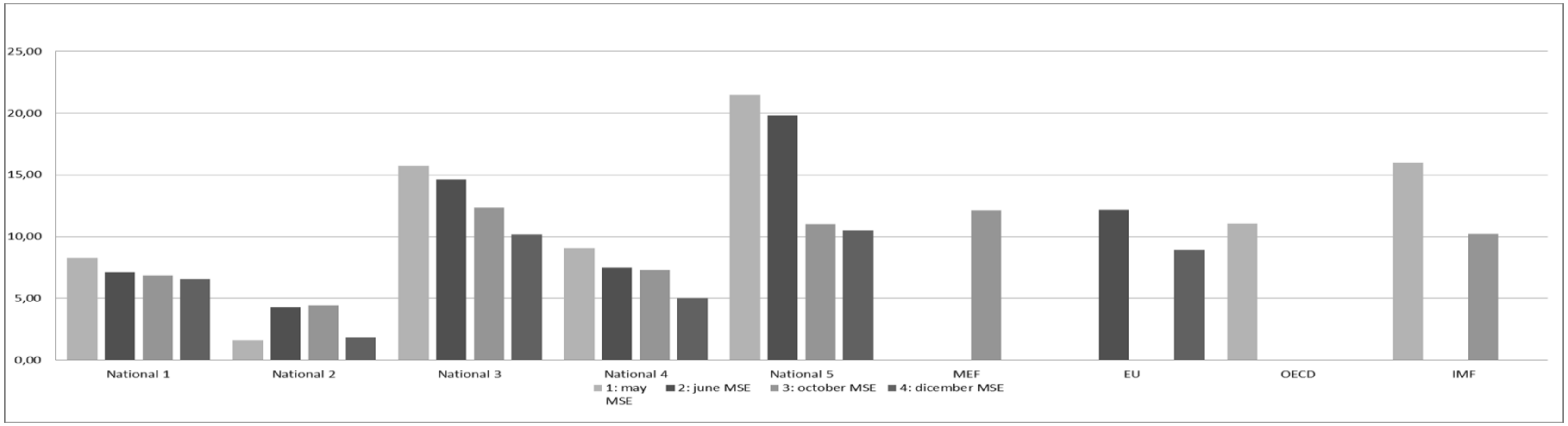


Figure II.6. Mean squared errors of year ahead forecast

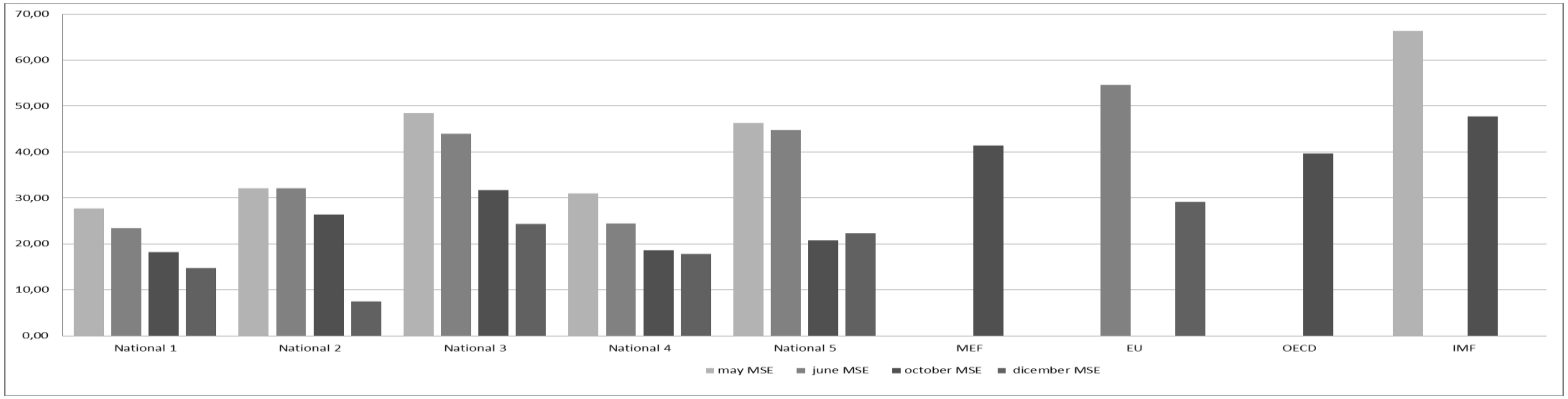


Figure II.7. Root mean squared error of current year forecast

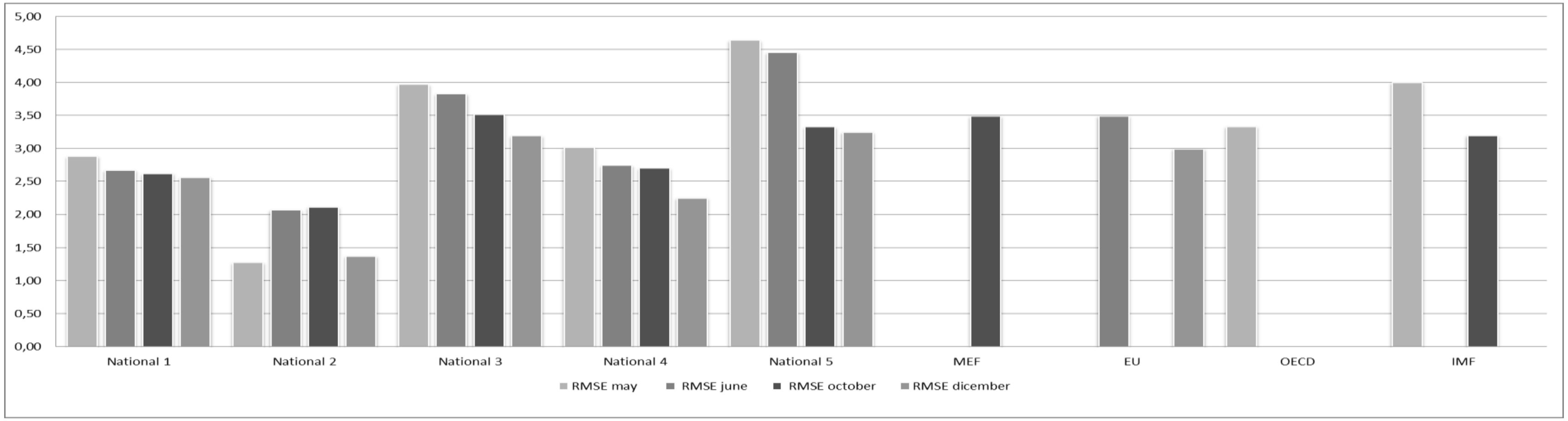


Figure II.8. Root mean squared error of year ahead forecast

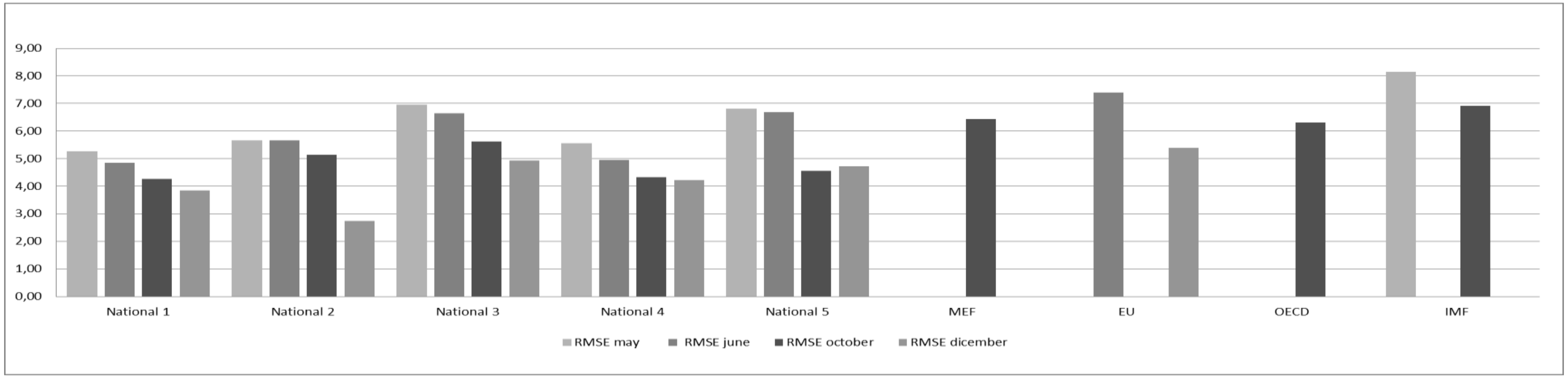


Figure II.9. Mean absolute deviation of current year forecast

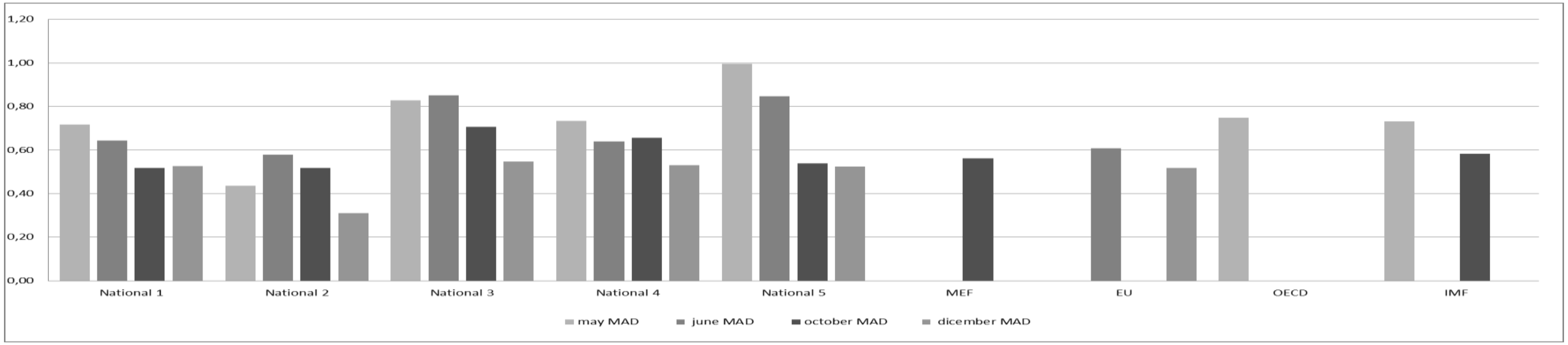


Figure II.10. Mean absolute deviation of year ahead forecast

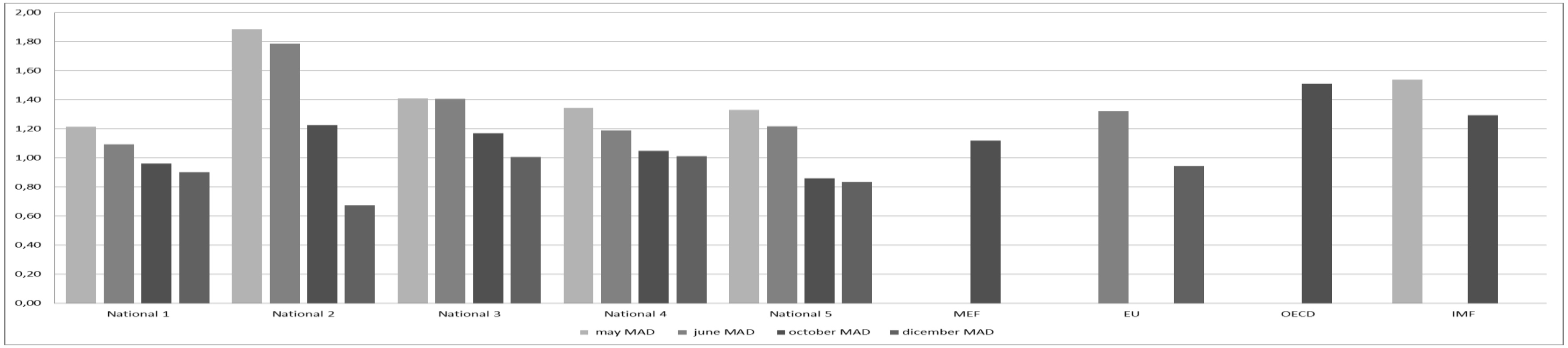


Figure II.11. Theil Index with naïve model of current year forecast

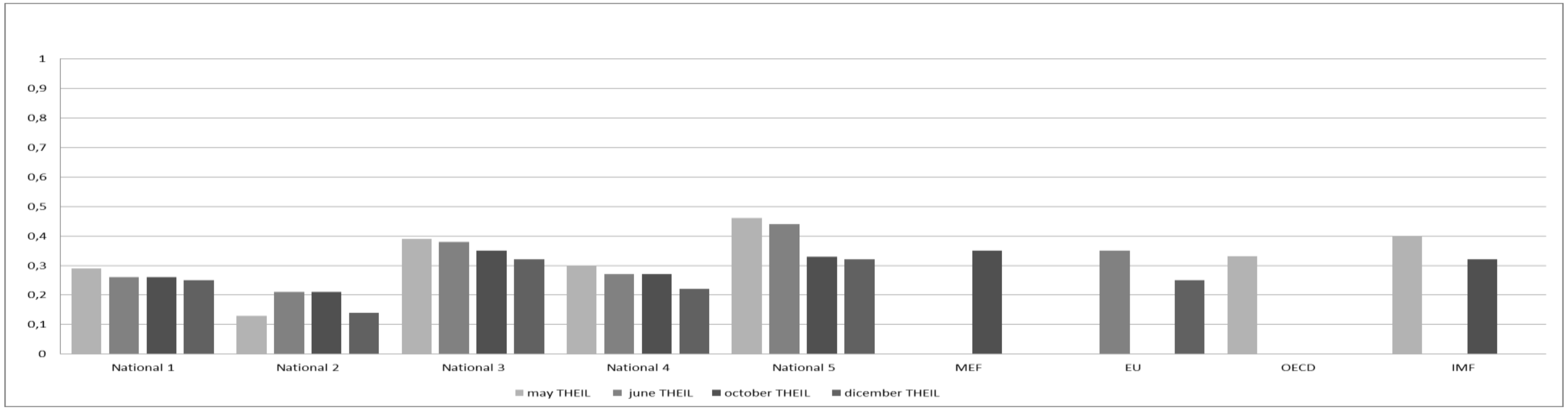


Figure II.12. Theil Index with naïve model of year ahead forecast

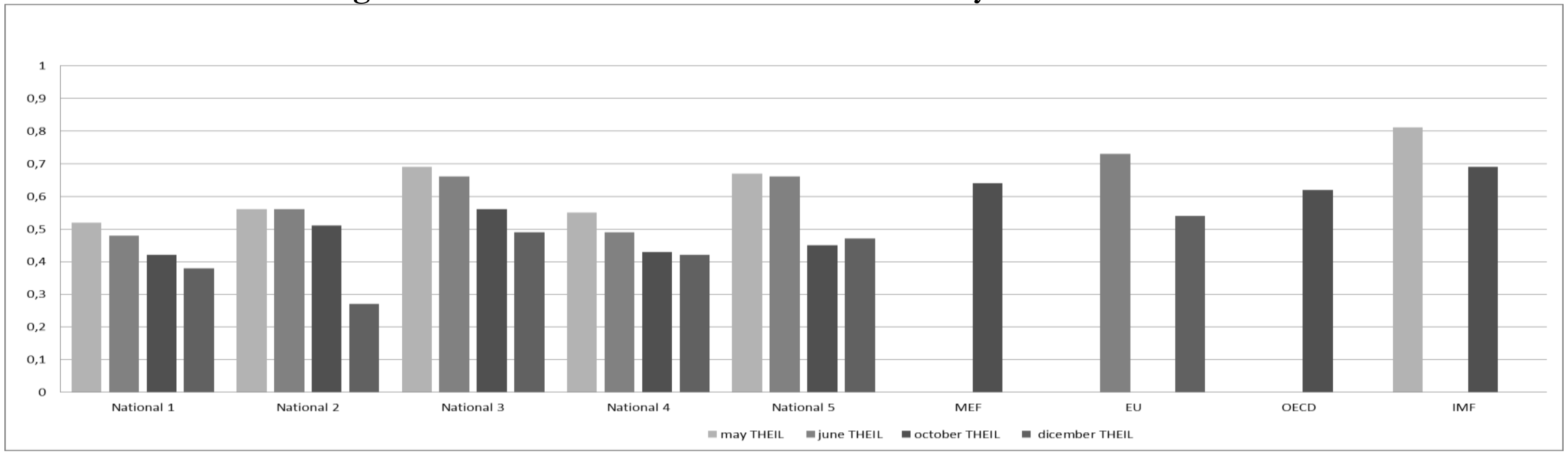


Figure II.13. Diebold and Mariano test – rank of current year forecast

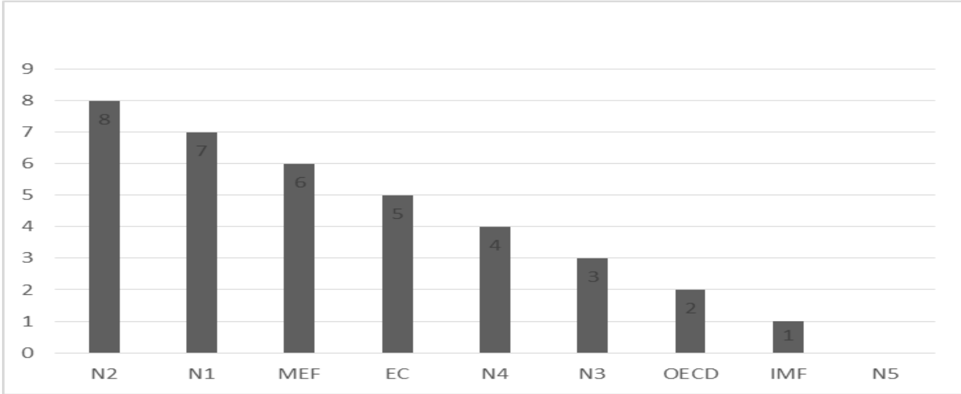
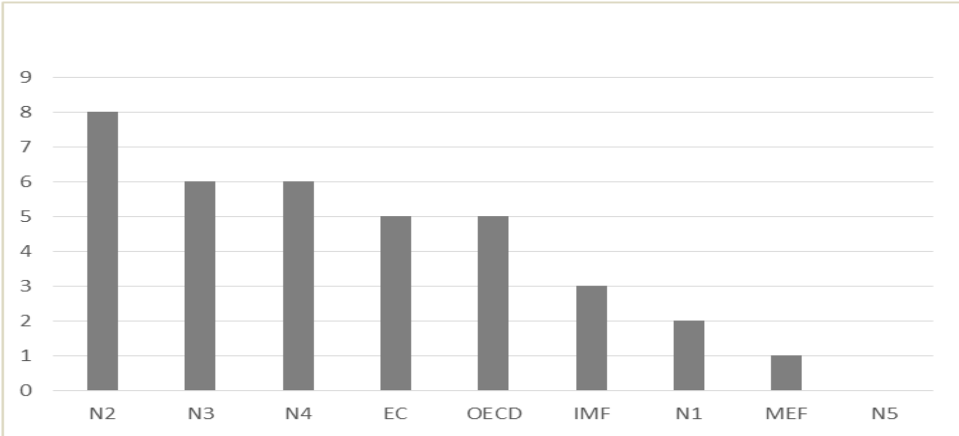


Figure II.14. Diebold and Mariano test – rank of year ahead forecast



Note: rank indicates the frequency with which each forecast is better than the others.
For instance: in current year and year ahead, National 2 is better of the other eight times.

The obtained results are consistent with most of the existing literature (Keereman, 1999, Artis and Marcellino 1998, 2001). A comparison of MSE, MAD, and RMSE for current year and year ahead forecasts indicates that the former are smaller than the latter for all agencies. This is not a surprise, as developments in the year are easier to forecast than events to come in the next year. For this reason too, the accuracy of forecasts improves as the year goes on. It can be observed, as a whole, that the accuracy of the forecasters improves from May to December, which is understandable given the fact that more complete information about public finances is available in December than in May. The exception is forecaster N_2, for whom the forecast in May is more accurate than for June. This may be due to the factoring in of changes in the legislature that normally occur at the end of April and May, as is shown in Table II.2. It is possible that N_2 gives more weight to political factors than the other forecasters when they make their public finance predictions.

The errors $e_{f,t}$ as shown by ME (Figure II.3.) show that all forecasters underpredict on average except N_2, N_1, N_5, and the IMF. With regard to MSE (Figure II.5.), RMSE (Figure II.7.), and MAD (Figure II.9.) the best performance is by N_2 and December is the month when forecasts tend to be the most accurate. For $e_{f,t+1}$ in Mean (Figure II.4.), all forecasters underpredict on average except for N_2 and the EU in October, and the IMF in June and October. The values of the MSE (Figure II.6.), RMSE (Figure II.8.), and MAD (Figure II.10.) show that in December, N_2 is the best forecaster, followed by N_1 for the months of October, June, and May. On all accounts – with ME, MSE, MAD and RMSE – the best performance is from private sector forecasters. It can be observed that private sector forecasts are more accurate than any of the national and international forecasters. Indeed, there is a sizable cross-forecaster variation. For every month $e_{f,t}$ and $e_{f,t+1}$, the best performance is by N_2. Figures II.11 and II.12 show the Theil index's results for current year and year ahead forecasts. Following the work of Artis and Marcellino (1998, 2001), I compare the forecasts of all forecasters with those of a naïve model. I compute the RMSE for the naïve model and compare this with the RMSE for the others forecasters. The Theil statistic is used to compare whether the forecasters are more accurate than the forecast of the naïve model. If the Theil values are smaller than one, this indicates that the forecasters outperform the naive model forecast on the basis of the RMSE. The results indicate that the RMSE of all forecasters are better than the naïve model. These results are similar to those shown in Artis and Marcellino (1998) and Keereman (1999), who find that in general the forecast errors made by forecasters are smaller than those obtained with a naïve model, showing at least some value of budget forecasting. Finally, Figure II.13 and II.14 show the Diebold and Mariano test rank of the numbers of the performances for current year and year ahead forecasts. According to the results from the Diebold and Mariano test, the private sector has the highest accuracy. In particular, the best performance is registered by N_2 for current year and for year ahead. The second and third best position also is filled by national agencies. Note that for year ahead, N_3 replaces N_1, and N_4 replaces MEF for second and third place, respectively. The international agencies have the fourth best position both for current year and year ahead.

II.5. Bias in forecasts

II.5.1. Methodology

A forecast is considered optimal when it meets certain properties, as discussed in Timmermann (2007). In particular, forecasts should be unbiased in the errors and have no serial correlation. In the literature, whenever these two properties are present, a forecast is called weakly efficient. In this section I consider all forecasts for variables $d_{f,t}$ and $d_{f,t+1}$ provided by each agency and for the selected months.

To achieve this, following the works of Artis and Marcellino (1998, 2001) and Keerman (1999), I compute a model where I formally analyze the unbiasedness and serial correlation of the forecasts by the following equations:

$$d_t=\alpha_0+\alpha_1 d_{f,t}+\mu \tag{II.9}$$

$$d_{t+1}=\alpha_0+\alpha_1 d_{f,t+1}+\mu \tag{II.10}$$

I test unbiasedness with a Wald test and check if the coefficient parameters comply with the following null hypotheses:

$$\alpha_0=0 \text{ and } \alpha_1=1 \tag{II.11}$$

The term μ is an error term that under the null hypothesis of unbiasedness coincides with the forecast error (Clements and Hendry, 1997).

Holden and Peel (1990) also showed that this condition is sufficient but not necessary for unbiasedness and suggested that the condition $\beta_0 = 0$ should be included in the regression:

$$e_{f,t} = \beta_0 + \nu_h \qquad \text{(II.12)}$$

$$e_{f,t+1} = \beta_0 + \nu_h \qquad \text{(II.13)}$$

where $e_{f,t}$ and $e_{f,t+1}$ are the error terms and υ_h is the demeaned forecast error. Weak efficiency further requires that the forecast errors be uncorrelated across time (Clements and Hendry, 1997).

II.5.2. Results

The results of testing efficiency, ubiasedness, and serial correlation are shown in Table II.2 for the agencies by month. Table II.3 summarizes the results for the agencies individually. Starting with the agencies taken as a whole, Table II.2 shows the results of the link between the actual data and the forecasts. For $d_{f,t+1}$and $d_{f,t}$, the model seems to indicate a strict link between these variables and the actual data. Table II.2, also for $e_{f,t}$ and $e_{f,t+1}$, seems to show a significant link between these two variables and the real data. With regard to efficiency, the results show that a pvalue under 0.05 indicates that the null hypothesis is rejected for all the forecasters analysed and for each month. This means that the forecasts provided are inefficient in all the samples. I also test the unbiasedness and the results show absence of bias when taking the sample as a whole, as each forecast has a pvalue above 0.05. Furthermore, I perform the Lagrange Multiplier (LM) test for lack of first and second order autocorrelation in the forecast errors. The results show that for each forecast, the hypothesis of autocorrelation is not rejected, with the exception of N_2 for December's deficit forecast and deficit error for the current year. Also, to evaluate if the coefficients are jointly significant I perform the Wald test. This test is used to test the sufficient condition on the hypothesis that the coefficient of (II.8) and (II.9) are jointly $\alpha_0=0$ and $\alpha_1=1$. The main result of the Wald test is that for each agency considering the high probability value, the null hypothesis cannot be rejected in the entire sample, with some exceptions (N_1 for June deficit forecast year ahead, N_2 for June deficit forecast current year and MEF in October forecast errors year ahead). In the case of both unbiasedness and uncorrelation, the forecasts and the forecast error are good parameters to explain the real data.

Table II.2. Unbiasedness and efficiency. LM test and Wald test for each agency taken individually

May					**June**					**October**					**December**				
N_1	**Bias**		**LM**	**WT**	**N_1**	**Bias**		**LM**	**WT**	**N_1**	**Bias**		**LM**	**WT**	**N_1**	**Bias**		**LM**	**WT**
	α_0	α_1				α_0	α_1				α_0	α_1				α_0	α_1		
Deficit Forecast current year	-0.60 (0.34)	0.86 (0.00)	0.27	0.62	Deficit Forecast current year	-0.51 (0.42)	0.91 (0.00)	0.34	0.66	Deficit Forecast current year	0.43 (0.34)	0.98 (0.00)	0.81	0.07	Deficit Forecast current year	-0.29 (0.49)	0.86 (0.00)	0.49	0.26
Deficit Forecast year ahead	-1.17 (0.12)	0.86 (0.00)	0.61	0.12	Deficit Forecast year ahead	-1,02 (0.10)	0.87 (0.00)	0.69	0.00 **	Deficit Forecast year ahead	0,11 (0.89)	1 (0.00)	0.81	0.67	Deficit Forecast year ahead	-0.22 (0.80)	1 (0.00)	0.89	0.57
Forecast error current year	-0.083 (0.76)		0.44		Forecast error current year	-0.21 (0.45)		0.34		Forecast error current year	0.48 (0.01)		0.82		Forecast error current year	0.21 (0.34)		0.88	
Forecast error year ahead	-0.47 (0.34)		0.82		Forecast error year ahead	-0.52 (0.29)		0.83		Forecast error year ahead	-0.007 (0.98)		0.09		Forecast error year ahead	-0.26 (0.43)		0.90	

May					**June**					**October**					**December**				
N_2	**Bias**		**LM**	**WT**	**N_2**	**Bias**		**LM**	**WT**	**N_2**	**Bias**		**LM**	**WT**	**N_2**	**Bias**		**LM**	**WT**
	α_0	α_1				α_0	α_1				α_0	α_1				α_0	α_1		
Deficit Forecast current year	0.28 (0.46)	1 (0.00)	0.43	0.36	Deficit Forecast current year	0.68 (0.21)	1 (0.00)	0.04	0.00 **	Deficit Forecast current year	-0.21 (0.63)	0.97 (0.00)	0.20	0.85	Deficit Forecast current year	-0.02 (0.93)	0.98 (0.00)	0.00 **	0.94
Deficit Forecast year ahead	-0.32 (0.56)	0.91 (0.00)	0.81	0.44	Deficit Forecast year ahead	0.16 (0.82)	0.98 (0.00)	0.28	0.82	Deficit Forecast year ahead	-0.59 (0.38)	0.89 (0.00)	0.34	0.65	Deficit Forecast year ahead	-0.53 (0.39)	1 (0.00)	0.65	0.13
Forecast error current year	0.27 (0.13)		0.34		Forecast error current year	0.45 (0.07)		0.10		Forecast error current year	-0.11 (0.60)		0.20		Forecast error current year	0.03 (0.80)		0.00 **	
Forecast error year ahead	1.10 (0.18)		0.81		Forecast error year ahead	1.10 (0.11)		0.95		Forecast error year ahead	0.79 (0.12)		0.65		Forecast error year ahead	0.17 (0.51)		0.79	

May				
N_3	**Bias**		**LM**	**WT**
	α_0	α_1		
Deficit Forecast current year	-0.66 (0.14)	0.87 (0.00)	0.72	0.32
Deficit Forecast year ahead	-0.68 (0.14)	0.94 (0.00)	0.59	0.16
Forecast error current year	-0.16 (0.53)		0.59	
Forecast error year ahead	-0.19 (0.66)		0.29	

June				
N_3	**Bias**		**LM**	**WT**
	α_0	α_1		
Deficit Forecast current year	-0.39 (0.43)	0.90 (0.00)	0.99	0.64
Deficit Forecast year ahead	-0.66 (0.14)	0.92 (0.00)	0.87	0.49
Forecast error current year	0.00 (0.99)		0.99	
Forecast error year ahead	0.00 (0.99)		0.29	

October				
N_3	**Bias**		**LM**	**WT**
	α_0	α_1		
Deficit Forecast current year	-0.47 (0.51)	0.83 (0.00)	0.95	0.40
Deficit Forecast year ahead	-0.03 (0.95)	1 (0.00)	0.73	0.84
Forecast error current year	-0.02 (0.92)		0.72	
Forecast error year ahead	-0.08 (0.84)		0.07	

December				
N_3	**Bias**		**LM**	**WT**
	α_0			
Deficit Forecast current year	-0.51 (0.18)	0.91 (0.00)	0.66	0.39
Deficit Forecast year ahead	-0.28 (0.54)	1 (0.00)	0.83	0.50
Forecast error current year	-0.2 (0.34)		0.67	
Forecast error year ahead	-0.27 (0.42)		0.31	

May					June					October					December				
N_4	**Bias**		**LM**	**WT**	**N_4**	**Bias**		**LM**	**WT**	**N_4**	**Bias**		**LM**	**WT**	**N_4**	**Bias**		**LM**	**WT**
	α_0	α_1				α_0	α_1				α_0	α_1				α_0	α_1		
Deficit Forecast current year	-0.68 (0.05)	0.86 (0.00)	0.95	0.14	Deficit Forecast current year	-0.83 (0.02)	0.86 (0.00)	0.95	0.06	Deficit Forecast current year	-0.78 (0.10)	0.87 (0.00)	0.59	0.25	Deficit Forecast current year	-0.57 (0.15)	0.89 (0.00)	0.43	0.29
Deficit Forecast year ahead	-0.92 (0.03)	0.89 (0.00)	0.99	0.07	Deficit Forecast year ahead	-0.90 (0.06)	0.91 (0.00)	0.77	0.09	Deficit Forecast year ahead	-0.89 (0.09)	0.91 (0.00)	0.74	0.16	Deficit Forecast year ahead	-0.82 (0.15)	0.95 (0.00)	0.40	0.14
Forecast error current year	-0.17 (0.48)		0.53		Forecast error current year	-0.23 (0.28)		0.57		Forecast error current year	-0.23 (0.41)		0.49		Forecast error current year	-0.07 (0.74)		0.27	
Forecast error year ahead	0.032 (0.94)		0.17		Forecast error year ahead	-0.04 (0.92)		0.27		Forecast error year ahead	-0.52 (0.23)		0.54		Forecast error year ahead	-0.23 (0.55)		0.33	

May					**June**					**October**					**December**				
N_5	**Bias**		**LM**	**WT**	**N_5**	**Bias**		**LM**	**WT**	**N_5**	**Bias**		**LM**	**WT**	**N_5**	**Bias**		**LM**	**WT**
	α_0	α_1				α_0	α_1				α_0	α_1				α_0	α_1		
Deficit Forecast current year	-0.89 (0.09)	0.78 (0.00)	0.92	0.17	Deficit Forecast current year	-0.67 (0.14)	0.84 (0.00)	0.75	0.25	Deficit Forecast current year	-0.27 (0.50)	0.92 (0.00)	0.61	0.58	Deficit Forecast current year	-0.29 (0.42)	0.91 (0.00)	0.42	0.41
Deficit Forecast year ahead	-0.92 (0.10)	0.83 (0.00)		0.24	Deficit Forecast year ahead	-0.60 (0.21)	0.96 (0.00)	0.67	0.18	Deficit Forecast year ahead	-0.79 (0.20)	0.94 (0.00)	0.99	0.19	Deficit Forecast year ahead	-0.69 (0.20)	0.93 (0.00)	0.99	0.26
Forecast error current year	-0.02 (0.93)		0.86		Forecast error current year	-0.03 (0.89)		0.65		Forecast error current year	0.07 (0.70)		0.69		Forecast error current year	0.09 (0.61)		0.57	
Forecast error year ahead	0.004 (0.99)		0.45		Forecast error year ahead	-0.02 (0.95)		0.50		Forecast error year ahead	0.10 (0.71)		0.55		Forecast error year ahead	-0.04 (0.87)		0.73	

May					June					October					December				
EC	**Bias**		**LM**	**WT**	**OECD**	**Bias**		**LM**	**WT**	**MEF**	**Bias**		**LM**	**WT**	**OECD**	**Bias**		**LM**	**WT**
	α_0	α_1				α_0	α_1				α_0	α_1				α_0	α_1		
Deficit Forecast current year	-0.95 (0.04)	0.85 (0.00)	0.83	0.11	Deficit Forecast current year	-0.53 (0.10)	0.88 (0.00)	0.24	0.17	Deficit Forecast current year	-0.33 (0.30)	0.94 (0.00)	0.54	0.58	Deficit Forecast current year	-0.62 (0.06)	0.90 (0.00)	0.28	0.16
Deficit Forecast year ahead					Deficit Forecast year ahead	-0.32 (0.43)	0.97 (0.00)	0.24	0.60	Deficit Forecast year ahead	-0.27 (0.50)	0.95 (0.00)	0.25	0.01 **	Deficit Forecast year ahead	-0.76 (0.08)	0.96 (0.00)	0.89	0.04 **
Forecast error current year	-0.32 (0.25)		0.48		Forecast error current year	-0.01 (0.93)		0.08		Forecast error current year	-0.11 (0.51)		0.44		Forecast error current year	0.18 (0.60)		0.85	
Forecast error year ahead					Forecast error year ahead	0.23 (0.54)		0.46		Forecast error year ahead	-0.19 (0.53)		0.11		Forecast error year ahead	-0.19 (0.48)		0.94	

May					June					October					December				
IMF	**Bias**		**LM**	**WT**						**MEF**	**Bias**		**LM**	**WT**					
	α_0	α_1									α_0	α_1							
Deficit Forecast current year	-0.77 (0.03)	0.83 (0.00)	0.32	0.06						Deficit Forecast current year									
Deficit Forecast year ahead	-0.92 (0.03)	0.89 (0.00)	0.99	0.07						Deficit Forecast year ahead	-1.06 (0.05)	0.83 (0.00)	0.72	0.15					
Forecast error current year	-0.03 (0.88)			0.05						Forecast error current year									
Forecast error year ahead	0.15 (0.72)			0.29						Forecast error year ahead	0.11 (0.81)		0.05						
										IMF	**Bias**		**LM**	**WT**					
											α_0	α_1							
										Deficit Forecast current year									
										Deficit Forecast year ahead	-1.06 (0.05)	0.83 (0.00)	0.72	0.15					
										Forecast error current year									
										Forecast error year ahead	0.11 (0.81)		0.05						

Note:** probability values indicate that the null hypothesis of Wald test and LM test has to be rejected at 5%.

Table II.3. Summary of efficiency and unbiasedness for the agencies by month

Month/ test	Test of efficiency	Test of unbiasedness	LM test	Wald test
December	Infefficent	Unbiased	Correlation deficit current and year ahead N_2	Coefficents jointly acepted
October	Infefficent	Bias in error current year N_1	Correlation error current year N_1	No jointly deficit year ahead MEF
June	Infefficent	Unbiased	Correlation deficit and error current year N_2 error current year OECD	No jointly deficit year ahead N_1 and and current year N_2
May	Infefficent	Unbiased	All forecasters	No jointly deficit year ahead OECD

Looking at Table II.3, I summarize the results considering the entire sample. I conclude that all agencies (national, private and public, and international) taken as a whole are good, but still not efficient at providing forecasts for current year and year ahead) and the forecasts could be improved using this information.

II.6. Conclusions and policy implications

In this Chapter, I focused on an analysis of forecast deficits, expressed as a ratio relative to GDP, made by international, national, and private agencies from the years 1992 to 2012 for the Italian budget deficit. I compared the current year forecast and the year ahead forecast and their relative forecast errors for each agency, depending on the month the forecast was released. I followed both a quantitative and a qualitative assessment of fiscal forecast errors. I applied the main tests of forecast accuracy.

The main result is that private agencies are more accurate in forecasting than others. The evidence shows, in general, a common prediction pattern for every agency for current year and year ahead forecasts. ME are in general small and negative, which implies that outturns are, on average, worse than projected. MAD and RMSE indicate that current year N_2 is the best forecaster for every month of the sample, while for year ahead N_2 is the best forecaster for December, and National 1 is the best forecaster for October, June, and May. Finally, the Theil test shows that all forecasters do better than a naïve forecast. I further tested for unbiasedness and serial correlation to show that each agency (with some exceptions) makes forecasts that are efficient and unbiased. When taken individually, the forecast data provided by the private agencies are the most accurate – even if they present weak efficiency conditions both for current year and year ahead.

A principal motivation for the analysis is that, given the stronger economic governance and coordination at the EC level, policy makers' decisions will depend increasingly on the forecasts made during the European semester. Can these forecasts then be used for the EC Commission to make recommendations to its member states? With regard to policy implications, it would be useful to have as much information as possible about the economic and financial situation of a country before issuing ex-ante strategies at the EC level. This is where the use of fiscal forecasts enters the picture. In particular, as many authors (Batchelor 2001, Abreu 2011, Frankel and Schreger 2011, 2013, Merola and Perez, 2013) argue, the budget-making process could possibly be improved by using the private-sector forecasts. Indeed, results show that throughout the year, a monitoring of their fiscal forecasts could be useful for the member states and for the EC when considering their high level of accuracy. It is also important to take into account forecast data that is provided in December of the previous year, because the year ahead forecast is a significant indicator of the future behaviour of that variable and would allow for better fiscal strategies to be created for the upcoming year.

II.7. References

Abreu, L. (2011). *International Organizations' vs. private analysts' forecasts: an evaluation* (Banco de Portugal working paper No. 20).

Artis, M., & Marcellino, M. (1998). *Fiscal solvency and fiscal Forecasting in Europe* (CEPR Discussion Paper No. 1836).

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal, 4,* 20-36.

Batchelor, R., (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics, 33,* 225 – 35.

Clements, M.P., & Hendry, D.F. (1997). Intercept corrections and structural change. *Journal of Applied Econometrics, 11*, 475-494.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253-263.

Favero, C., & Marcellino, M. (2005). *Modelling and forecasting fiscal variables for the Euro Area* (Bocconi University Working Paper No. 298).

Frankel, J. A., & Schreger, J. (2011). *Over-optimism in forecasts by official budget agencies and its implications* (NBER Working Paper No. 17239).

Frankel, J. A., & Schreger, J. (2013). Bias in official fiscal forecasts: Can private forecasts help? mimeo, Harvard.

Holden, K., & Peel, D.A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School, 58*(2), 120 -127.

Keereman, F. (1999). The track record of the Commission forecasts (EC, Directorate-General for Economic and Financial Affairs Economic Paper No. 137).

Merola, R., & Pérez, J. J. (2013). Fiscal forecast errors: Governments versus independent agencies? *European Journal of Political Economy, 32*(C), 285-299.

Theil, H. (1971). *Applied economic forecasting*. North-Holland Publishing Company.

Timmermann, A. (2007). An evaluation of the World Economic Outlook forecasts. *IMF Staff Papers, 54*(1), 1–33.

ANNEX II.1 Measurement of the forecast errors

Table A.II.1.1 Diebold and Mariano test for current year

DMT	**N_1**									
N_1	MSE	**N_2**								
N_2	0.006 N_2 (0.00	MSE	**N_3**							
N_3	0.007 N_1 (0.01	0.02 N_2 -0.66	MSE	**N_4**						
N_4	-0.32 N_1 0	-0.1 N_2 -0.001	0.14 N_4 -0.26	MSE	**N_5**					
N_5	-0.33 N_1 0	-0.25 N_2 -0.004	-0.11 N_3 -0.23	-0.38 N_4 0	MSE	**MEF**				
MEF	-0.01 N_1 -0.94	0.08 MEF -0.12	0.29 MEF -0.01	0.04 MEF -0.34	0.6 MEF -0.6	MSE	**OECD**			
OECD	-0.15 N_1 -0.2	-0.04 N2 -0.89	0.26 OECD -0.06	0.4 OECD -0.54	0.56 OECD -0.04	0.52 MEF -0.23	MSE		**EC**	
EC	-0.16 N_1 -0.2	-0.41 N2 -0.26	-0.04 N_3 -0.46	-0.19 N_4 -0.22	0.13 EC -0.1	-'0.20 MEF 0	-0.07 OECD -0.23		MSE	**IMF**
IMF	-0.33 N_1 -0.16	-0.17 N2 -0.05	0.06 IMF -0.72	-0.36 N_4 0	0.38 IMF -0.15	-'0.17 MEF 0	-0.171 OECD -0.65		-0.16 EC 0	

Note(a): * denote that the horizontal subject is the best of the vertical subjects , ** denote that the vertical subject is the best of horizontal subjects, = indicate that horizontal and vertical subjects predict in the same way. Note (b): N_1= national 1, N_2= national 2, N_3= national 3, N_4= national , N_5= national 5. Note(c): "MSE" is mean square error.

Table A.II.1.2. Diebold and Mariano test for year ahead

DMT	**N_1**								
N_1	MSE	**N_2**							
N_2	0.33 N_2 (0.02	MSE	**N_3**						
N_3	0.59 N_3 (0.28	-0.28 N_2 0	MSE	**N_4**					
N_4	0.38 N_4 -0.53	-0.51 N_2 0	-0.03 N_3 -0.99	MSE	**N_5**				
N_5	-0.06 N_1 -0.85	-0.1 N_2 -0.07	-0.41 N_3 0	-0.17 N_4 -0.03	MSE	**MEF**			
MEF	-0.19 N_1 -0.64	-0.65 N_2 0	-0.62 N_3 0	-0.42 N_4 0	-0.19 MEF -0.38	MSE	**OECD**		
OECD	0.31 OECD 0	-0.32 N_2 -0.24	0.03 OECD -0.88	-0.07 N_4 -0.35	0.31 OECD -0.14	-0.5 EC -0.17	MSE	**EC**	
EC	0.03 EC -0.66	-0.03 N_2 -0.75	-0.26 N_3 -0.64	-0.25 N_4 -0.41	0.14 EC -0.7	0.51 EC -0.23	-0.34 OECD -0.16	MSE	**IMF**
IMF	0.09 IMF (0,46)	-0.32 N_2 (0,00)	-0.4 N_3 (0,06)	-0.23 N_4 (0,03)	0.03 IMF (0,74)	0.23 EC (0,40)	-0.27 OECD (0,28)	0.16 IMF (0,87)	

Note(a): * denote that the horizontal subject is the best of vertical subjects, ** denote that the vertical subject is the best of horizontal subjects, = indicate that horizontal and vertical subjects predict in the same way. Note (b): N_1= national 1, N_2= national 2, N_3= national 3, N_4= national, N_5= national 5. Note(c): "MSE" is mean square error.

Chapter III.

Combine to compete: improving fiscal forecast accuracy over time[8]

III.1. Combining deficit forecasts to improve accuracy

As mentioned in the previous Chapters, budget forecasts have become increasingly important as fiscal management tools that influence the expectations of bond markets and the public at large. However, the inherent difficulty in projecting macroeconomic variables – together with political bias – impede the accuracy of budget forecasts. In this Chapter I focus on improving accuracy by combining the forecasts on deficit/GDP of both private and public Italian agencies over the 1992-2012 period. Evidence tells us that budget forecasts have been a rather poor guide for correctly assessing the fiscal outlook, especially if they are based on government data. The projections often paint too rosy a picture of reality, and are consistently biased towards low deficits, especially when confronted with comparable predictions made by international institutions (Artis and Marcellino, 1998, 2001, Afonso et al., 2004). Projections of fiscal adjustments are usually pushed forward over time, and revised when the decision nears (Beetsma and Giuliodori, 2010). A large literature argues that this bias in prediction performance is the consequence of setting politically motivated targets rather than realistic economic projections (Fildes and Stekler, 2002, Jonung and Larch, 2006). Nevertheless, even the forecasting performance of private institutions and public agencies is not stellar, which casts doubt on the ability to forecast fiscal variables (Favero and Marcellino, 2005). This might

[8] I presented this Chapter at the AQR/IREA seminar at the Universitat de Barcelona; at the ISF 34th International Symposium on Forecasting in Rotterdam (Holland) in July 2014; at the PhD Worskshop at the Universitat de Barcelona in December 2014; and at the VI Workshop on Time Series Econometrics in Zaragoza (Spain) in April 2015. The Chapter was also accepted at the Workshop on Macroeconometrics DIW in Berlin (German); the Storep "Shifting Boundaries: Economics in the Crisis and the Challenge of Interdisciplinarity" conference in Torino (Italy); the 12th Edition of the ACDD Augustin Cournot Doctoral Days conference in Strasbourg (France); the 2nd Conference of the International Association for Applied Econometrics (IAAE 2015) at the University of Macedonia in Thessaloniki (Greece); and Time Series-ITISE 2015 at the University of Granada (Spain) with four positive referees. It is published as a working paper of the Universitat de Barcelona in coauthorship with Peter Claeys and available on-line via the Repec/Ideas and SSRN platforms.

be the result of a lack of attention by private agencies, as budget forecasting has not been a priority for them. Public agencies like the OECD, IMF, or the EC have been facing significant information problems, in spite of the more advanced economic models used for forecasting. Research into better practices for forecasting budget variables has not come to any conclusive findings. The bottom line of most applied work is that results depend on the forecasting procedure chosen, the consistency of macroeconomic and fiscal forecasts, the forecast horizon, and the level of disaggregation of fiscal forecasts. Efforts to improve data availability on fiscal accounts over the last decade have paid off as attempts to incorporate more detailed information (Onorante et al., 2010; Pedregal and Perez, 2010), and to apply more advanced econometric techniques (Asimakopolous et al., 2013) have led to marginal improvements in forecasting performance. Forecasting the budget deficit is still considered to be more of an art than a science. Fiscal forecasts may require more judgement and expertise than econometric or modelling techniques (Leal et al., 2008). If progress depends on better inside knowledge of the dark box of the budget process, then the ultimate consequence is that there may be as many forecasts as there are forecasters.

The objective of this Chapter is to improve forecast accuracy by exploiting the information contained in all the individual budget forecasts that were analysed in Chapter II. I do this by averaging forecasts from different sources in a variety of ways. I include simple as well as more advanced averaging techniques that account for past forecasting performance to compute a combined forecast, and then I check if the performance is robust in the face of changes over time.

The Chapter is structured as follows. I first review the methodology in section III.2; in section III.3, I analyse several techniques for combining forecasts, ways to evaluate and compare forecasts (over time), and the data. In section III.4, I discuss the tests to compare the combination of forecasts to other forecast models and then, in section III.5, I consider their evolution over time by testing the accuracy of fiscal forecasts, and their stability over time by using the fluctuation test developed by Giacomini and Rossi (2010). I do this measuring the out-of sample MSFE differences computed over rolling windows and testing the null hypothesis of zero MSFE differentials between the two competing models at each point in time in the forecast evaluation period, using the critical values provided by the table from Giacomini and Rossi (2010). In section III.5., I conduct some robustness checks. Section III.6. concludes the Chapter.

III.2. Methodology

A vast literature shows that the combination of various forecasts results in improved prediction performance (Clemen, 1989; Clemen and Winkler 1986, Clements and Hendry, 2003; Timmermann, 2006). Many authors have studied the pooling of forecasts. Zarnowitz (1967), for example, noted that the published averages of inflation and GNP growth forecasts were better than the individual forecasts themselves. Bates and Granger (1969) discovered that the simple average outperforms the forecasts taken individually. The idea was also to use the relative combination of variances and covariances to construct a weighted average of the forecasts that minimizes the mean square error of the combined forecasts. Likewise, Nelson (1972) and Cooper and Nelson (1975) showed that the combination of forecasts with ARIMA estimates produces a smaller error compared to the models alone. The suggested reasons for the better performance of ARMA models in their paper are the incapacity of econometric models to arrange structural changes in the economy. Granger and Newbold (1973) also start from a similar point in terms of forecast evaluation. Makridakis (1982, 1983, and 1989) studied a large variety of time series forecasting methods, which were applied to 1,001 different economic time series.

The forecast performance was measured using various error summary measures. Two different combining schemes were studied: both of these combinations performed well relative to the individual techniques, with the simple average having the better performance of the two. Clemen (1989) provided a very deep review of the methods used in combining and confirming these results. Clemen and Winker (1986) offer a philosophical approach to the idea of combination.

The reason for the improved performance is that single forecasts are the product of a specific forecasting model, which depends on specific econometric techniques and personal judgment - each of which have some idiosyncratic errors. Pooling many forecasts averages out these errors. Also, the empirical models used in forecasting are based on the assumption of stable relationships, but political events, crises, technological progress, etc. upset economic relations continuously. Combination levels out this instability (Pesaran et al., 2004).

Further, combining reduces the risk of forecast bias when there are many macroeconomic variables that are endogenous over the economic cycle. If forecasts are used as a proxies input for forecasting other variables, these proxies introduce a systematic measurement bias and reduce forecast accuracy. Finally, each forecasting model assumes a loss function by the forecaster. With changes in volatility of the economic variables used in the model, combining forecasts can produce more precise results. Thus the aim of such combinations is to make forecasting practices robust to the different types of uncertainty.

A combined forecast Y^*_{it+h} of n different forecasts of a variable Y at horizon h is of the general form:

$$Y^*_{t+h}=\alpha_t+\sum_{i=1}^{n}\beta_{i,t}Y_{i,t+h} \qquad \text{(III.1)}$$

A considerable amount of research has been undertaken to determine how best to choose the coefficients, α_t and β_{it}. Evidence suggests that the simple approach of averaging the individual predictions works well (Lupoletti and Webb, 1986; Clemen and Winkler, 1986; Clemen, 1989). In this case, β_{it} is equal to 1/n on all individual predictions. Alternatively, the geometric mean and harmonic mean and the median can be used as a summary. The simple average has often been found to be quite a robust forecast for a set of economic variables, suggesting that forecasters are on average right (Clemen, 1989).[9] More complex methods may further improve performance by attributing different weights $\beta_{i,t}$ to each forecaster. The typical way to do this is to give more weight to better performing forecasters, for example by attributing to each expert forecaster a weight that is inversely proportional to the predictor's Mean Square Forecast Error (MSFE). The proposal is to value more those forecasters that have a higher past average performance. However, past performance might not be a good guide to future performance: the resilience of a forecasting model to structural breaks distinguishes good from average forecasts. Recent performance is therefore more relevant for forecast evaluation than average historical performance. Stock and Watson (2003, 2004) discount past MSFE over a horizon *h* to attach greater weight to the recent predictive ability of each individual predictor. The weights in (III.1) are then given by the forecaster's

[9] Note that estimating the combination weights might induce uncertainty, especially when the sample size is small relative to the number of forecasts (Elliott, 2004).

MSFE compared to the overall MSFE, where past MSFE is discounted *h* periods back in time with a factor δ (III.2):

$$\beta_{i,t+h} = \frac{m^{-1}_{i,t}}{\sum_{i=1}^{N} m^{-1}_{i,t}} \quad \text{where } m_{i,t+h} = \sum_{s=t_0}^{t-h} \delta^{t-h-s} \left(Y^h_{s+h} - \hat{Y}^h_{i,s+h}\right)^2 \qquad \text{(III.2)}$$

As forecasters update their models quickly after bad performance, one should exclude out-dated versions of forecasting models. Stock and Watson (2004) propose to cut off information from all past performance after a relevant period of time. This corresponds to setting δ to 1, and reducing *h* to a short horizon, and gives the 'best recent' forecast. An alternative approach to computing the weights is to estimate the weights from a simple regression of the forecast on the different forecasts, as in (III.3):

$$Y^*_{t+h} = \alpha_t + \sum_{i=1}^{n} \beta_{i,t} Y_{i,t+h} \qquad \text{(III.3)}$$

This is nothing else than an extended version of the regression used for testing the unbiasedness and weak efficiency of the forecast, where for a single forecast, α_t should be zero and $\beta_{i,t}$ should be 1. The regression approach relaxes the assumption in (III.2) of unbiased and uncorrelated errors as the constant is not bound to be zero, and the weights do not sum to 1 (Lupoletti and Webb, 1986).

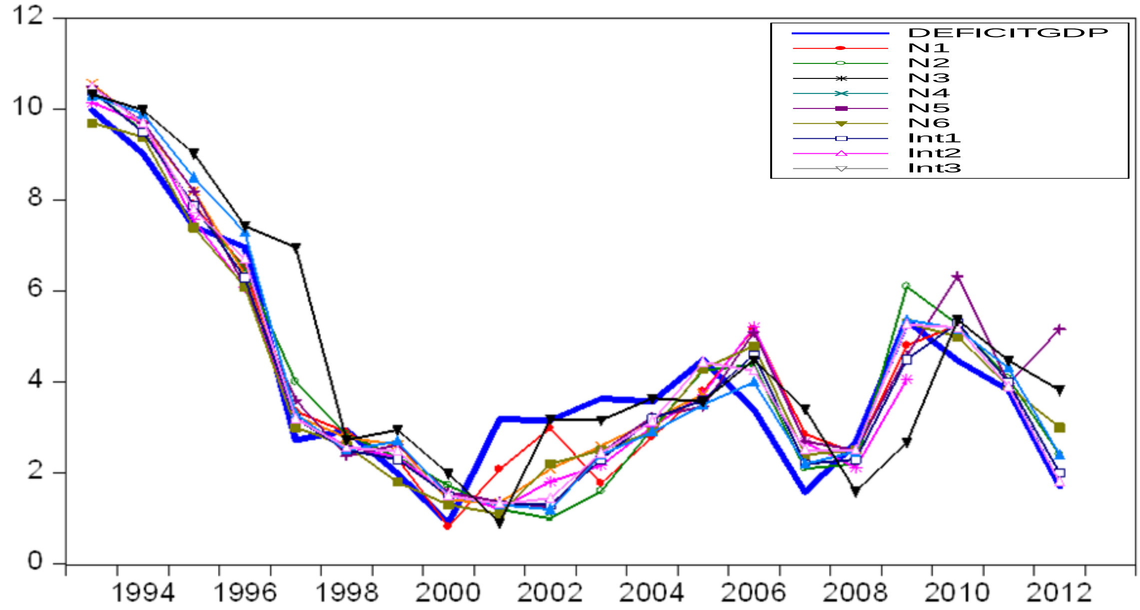
III.3. Combinations of fiscal deficit forecasts in Italy

In this Chapter, I consider the same database as in the previous one. Indeed, in this database I include the information from the months of May or June for public institutions (EC, OECD, IMF, and MEF) and May for the forecasters from the CEF database. In a few cases, some of the private forecasts were missing; in such cases I used the forecasts from April that year. In this Chapter, I also add a simple naive forecast to the nine forecasts, which is just the realised deficit ratio of last year. Figure III.1(a) shows a graph of the different current year forecasts over time, and compares them to the realised deficit to GDP ratio for that period. This deficit ratio d_t comes from the OECD Economic Outlook.[10] Figure III.1(b) does the same for the deficit forecasts one year ahead, provided by the same respondents.

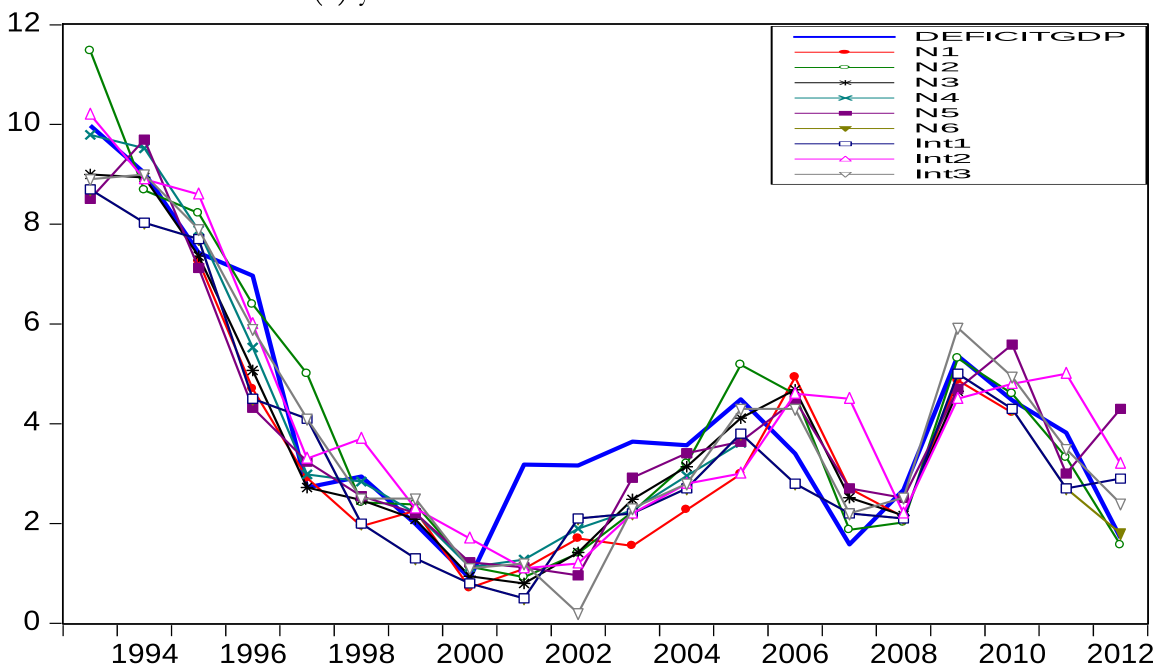
[10] Note that revisions to the deficit ratio are subject to adjustment for a couple of years after its first publication.

Figure III.1. Realised and forecast deficit ratio, sample 1992-2012

(a) current year forecast



(b) year ahead forecast



Note: N6 is MEF, Int 1 is EC, Int 2 is OECD and Int 3 is IMF

In both panels of Figure III.1, the forecasts broadly move in the same direction, but there is definitely more dispersion in the year ahead forecasts than in the current year forecasts. While the range of forecasts differs by no more than 1 per cent of GDP in the latter, the range increases to 3 per cent on average for the former. There are also considerable changes over time. Up to 2001, all forecasters agree on a quite fast consolidation, and this is certainly inspired by the Maastricht criteria. Afterwards, the forecast tends to become less accurate. The exception is the rapid rise of deficits during the Financial Crisis: all forecasters agreed that the deficit would become much larger. The consolidation in the following year is not as easily foreseen.

I have nine expert forecasts whose specific information sets underlying their forecasts are unobserved, so pooling the forecasts may add value. I compute 9 different combined forecasts. These include four simple combination models that average the different deficit forecasts (simple average, geometric average, harmonic average and median). I then compute three regression weight based combination models. The first one is based on the regression of the realised deficit ratio on all nine forecasts (weighted forecast combination, WFC).[11] Pesaran et al. (2004) show that including models with different degrees of adaptability to breaks outperforms the forecasts from alternative pooled forecasts. The naive model picks up any such changes in the following period, whereas the other forecasts may still deviate due to their dependence on past patterns.

One of the benefits of the CEF forecasts is that unlike other surveys, individual forecasts in the CEF should not suffer a bias owing to the release of strategic forecasts, as often happens for official forecasts released by governmental agencies (Ottaviani and Sorensen, 2006). CEF data are public, which prevents a participant from reproducing others' forecasts and also limits the possibility of herding (Trueman, 1994). Analysts are bound in their survey answers by their recommendations to their clients, and discrepancies between the survey and their private recommendations would be hard to justify (Keane and Runkle, 1990). In addition, and unlike other surveys, professional economists who participate in the CEF poll not only take a stance on the direction of the expected change of a macroeconomic variable, but also forecast the level of the macroeconomic variable. Evidence shows that CEF forecasts are less biased and

[11] For the detailed results of the regressions of the weighted forecast combinations, see Annex III.1.

more accurate than other surveys. One might reasonably expect then that private forecasters outperform the public forecasts too.[12]

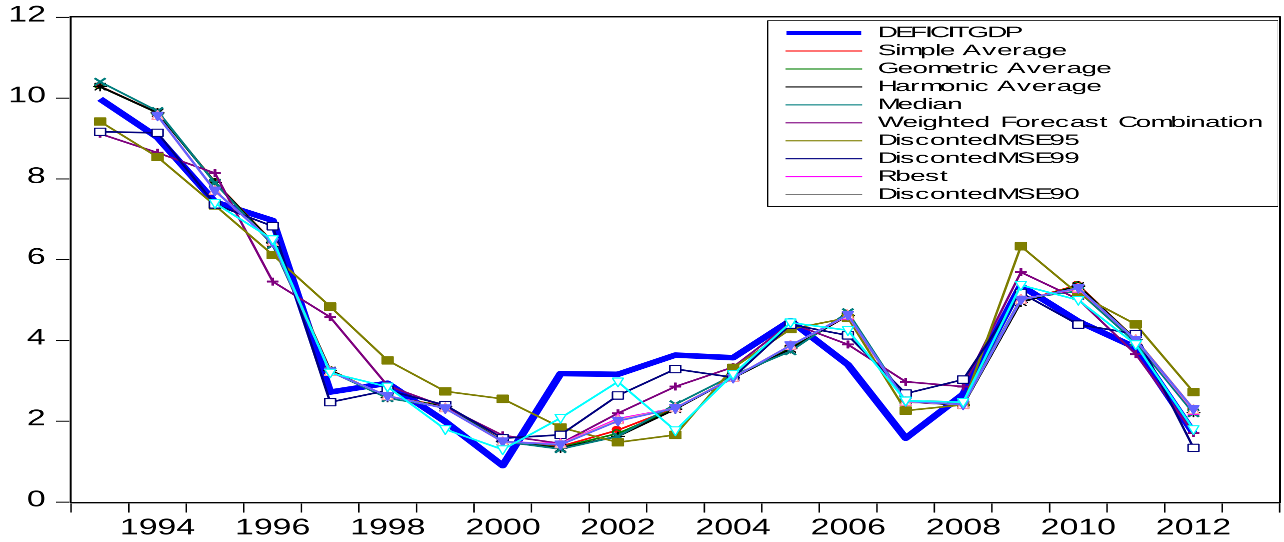
I next construct four forecast combinations that select only the best performing forecasters over recent periods. I apply the weights from (III.2) that are inversely proportional to the predictor's Mean Square Forecast Error (MSFE) relative to the realised deficit ratio. For the first three of these combinations, I discount past performance using a value of δ of 0.90, 0.95, and 0.99 (respectively: $disc_{90}$, $disc_{95}$ and $disc_{99}$). Alternatively, I cut off the time horizon after four years and look only at the recently best-performing forecasters (R_{best}).

Figure III.2 displays the realised deficit d_t together with the different combined forecasts. Panel (a) shows the current year forecast, and panel (b) the year ahead forecast. All combined forecasts track the deficit closely over the first part of the sample (up to 2001). Afterwards, there is a tendency to deviate from the deficit for a couple of years. Figure III.2 shows that most expert forecasts fail in the same direction at the moment of an unexpected break. A comparison of Figure III.2 to the original forecasts shows that combinations are less variable than the single forecasts. Panel (a) of figure III.2 suggests that all forecast combinations are equally good in tracking the realised deficit. In panel (b), the weighted forecast combination, as well as the R_{best} combination, are closest to the actual data in 2001-2002 when all the agencies tended to make large forecasting mistakes. Combination is unlikely to provide a substantial improvement over the best individual forecasts in such a setting.

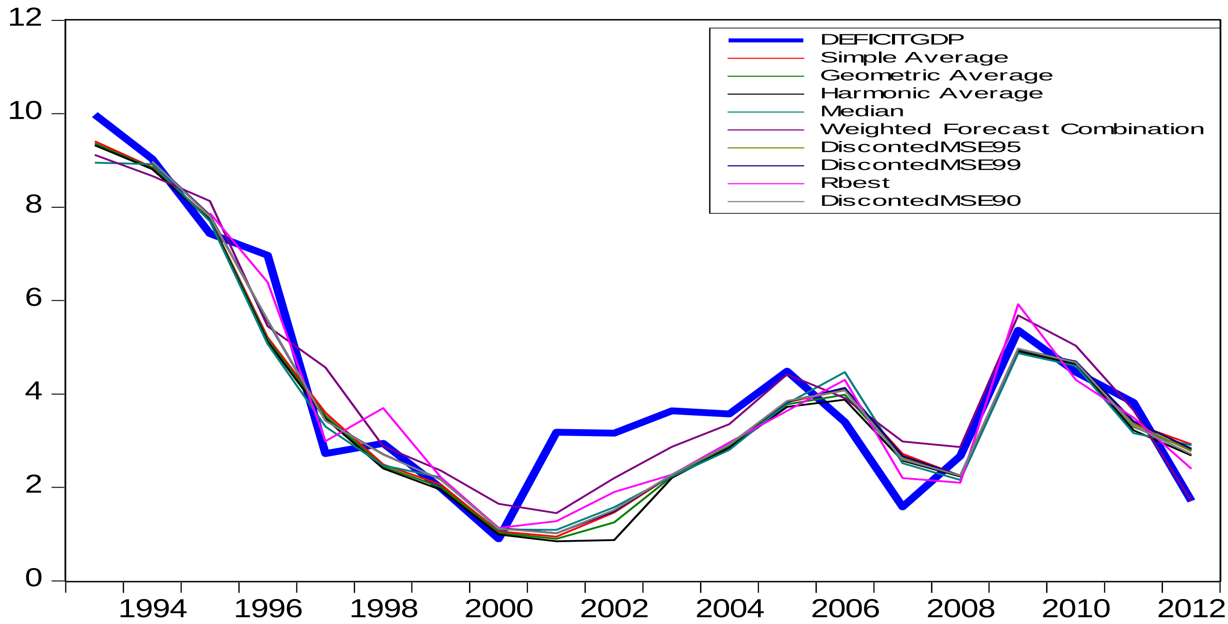
[12] Batchelor (2001) shows that CEF forecasts are less biased and more accurate in terms of mean absolute error and root mean square error than OECD and IMF forecasts for the variables: real GDP, consumer spending, business investment, industrial production, inflation and unemployment. Dovern and Weisser (2011) also find that the participants in the CEF poll provide rational and unbiased inflation and growth forecasts for the G7 countries.

Figure III.2. Realised deficit ratio and the combined forecasts.

(a) current year forecast



(b) year ahead forecast



III.4. Predictive accuracy of combined forecasts of fiscal deficit in Italy

III.4.1. Tests of predictive accuracy

An eyeball comparison of Figures III.1 and III.2 suggests that the combination forecasts likely outperform the original ones. I analyse if a linear combination of forecasts outperforms any individual forecast or an AR model. I apply standard tests and compute the RMSE, the MSE, the MAD, and the Theil test (1958) on each forecast, as compared to the realised deficit ratio that was described in the previous Chapter. In a second exercise, I apply the Diebold-Mariano (1995) test of predictive accuracy and compare each of the 20 single or combined forecasts with the combined forecast and with a simple naive model.

III.4.2. Accuracy of forecast errors

I see from Table III.1 that all forecasters – public, private, and combined – do much better than a simple naïve model would suggest, both for the current year and the year ahead forecast. This stands a bit in contrast with other results in the field that find that the naive model performs at least as well as public forecasts (Artis and Marcellino, 2001, Marcellino 2002, 2004) or simple time-series models (Favero and Marcellino, 2005). On the current year forecast, public forecasters generally outperform the private ones. The performance of the combined forecasts is a bit mixed as a result. Unsurprisingly, the simple combinations of private and public forecasts tend to do worse than the public forecasters. The overall combination (WFC) improves over the public forecasts as it puts less weight on the private forecasts, and aggregates the information of the public ones. The robustness to structural breaks explains why the R_{best} is more accurate than the discounted combination of forecasts. It is unsurprising that forecasting performance for deficits one year ahead is worse than for the current year. All forecasters – or any combination of them – do better than the simple naive model. Evidence is more mixed on the relative performance of private and public forecasters: the IMF or EC forecasts are ranked below those of the MEF, OECD, or any private forecaster.

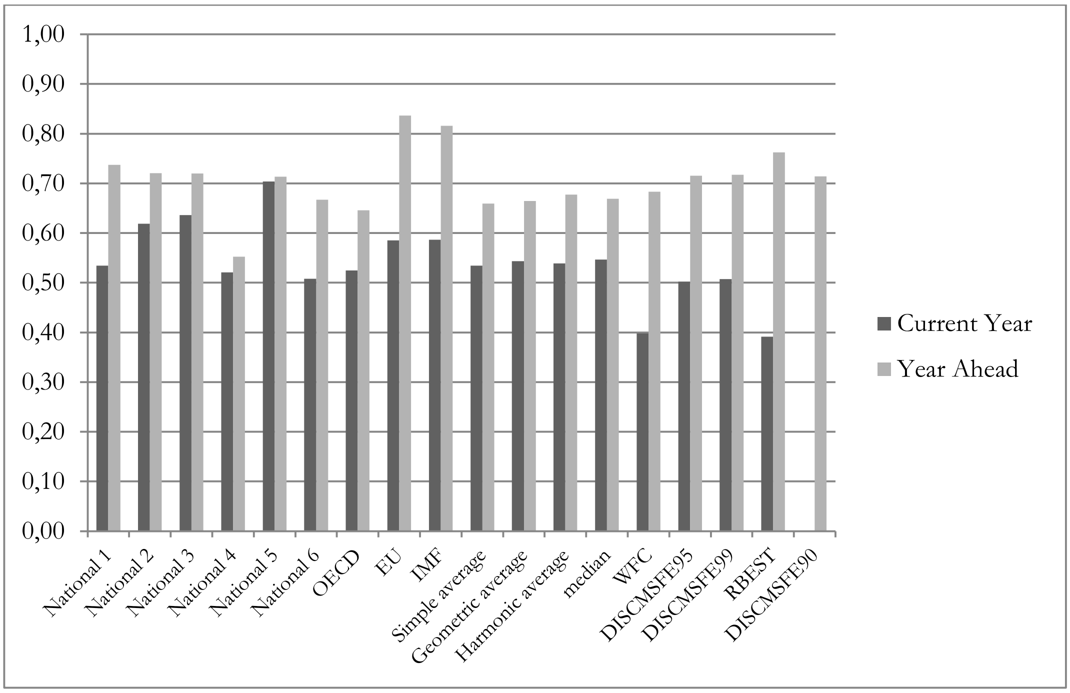
Table III.1. Accuracy test of single and combination forecasts

Forecast combination	Current year			Year ahead		
	RMSE	MAD	MSE	RMSE	MAD	MSE
N_1	3.73	0.72	1.93	7.15	1.39	2.67
N_2	4.10	0.74	2.02	6.99	1.68	2.64
N_3	4.22	0.84	2.05	6.98	1.34	2.64
N_4	3.45	0.74	1.86	5.36	1.25	2.31
N_5	4.66	0.97	2.16	6.92	1.22	2.63
MEF	3.37	0.60	1.84	6.47	1.17	2.54
OECD	3.48	0.66	1.86	6.27	1.09	2.50
EC	3.88	0.78	1.97	8.11	1.64	2.85
IMF	3.89	0.76	1.97	7.91	1.50	2.81
Simple mean	3.73	0.72	1.93	6.60	1.24	2.57
Harmonic mean	3.80	0.72	1.95	6.78	1.25	2.60
Geometric mean	3.77	0.72	1.94	6.65	1.25	2.58
Median	3.82	0.71	1.95	6.69	1.25	2.59
WFC	2.78	0.48	1.67	6.83	1.38	2.61
$disc_{90}$	- [(a)]	-	-	7.14	1.39	2.67
$disc_{95}$	3.51	0.69	1.87	7.16	1.39	2.68
$disc_{99}$	3.54	0.70	1.88	7.18	1.39	2.68
R_{best}	2.73	0.44	1.65	7.62	1.52	2.76
Naive	6.99	1.20	2.64	10.00	1.84	3.16

Notes: (a) not available as some expert forecasts are not available at a sufficiently long time horizon.

Figure III.2 already showed that forecasts deviate more from the realised deficit than from the current year forecast. As a result, the combinations do not provide much improvements over the single forecasts. Measured by the RMSE, forecaster N_4 actually beats all of the other forecasts. The reason for the relative underperformance of the combination must be that all forecasts are now prone to make mistakes due to structural breaks. In Figure III.3, the Theil test (1958) shows us the improvement in performance relative to the naïve model. As I computed already in Table III.1, any single public or private forecast, or a combination of them, does much better than the naive model. Unsurprisingly, the accuracy is always better for individual and combination models in the current year as compared to a forecast of the year ahead. The graph confirms that the R_{best} improves considerably on accuracy for current year predictions; this is not generally the case for the year ahead forecast. In fact, respondent N_4 produced a forecast that is about 10 per cent more accurate than the best combined forecast, which in this case is the simple average of all nine forecasts.

Figure III.3. Theil test of single and combination forecast



III.4.3. Diebold and Mariano test

The accuracy criteria show that the improvements in forecast performance from one forecast to another are often marginal, so they may not be significant. I therefore compare each public/private forecast to a combination forecast, and test its predictive performance. Table III.2 summarises the results of the DM test. The left part of the table shows the comparison of all combined forecasts to the private and public forecasts, and the right part shows the comparison between the different combined forecasts. Grey shaded cells indicate that the significant 'winner' of the contest is the combined forecast in the left column of the table. In contrast to most of the literature on forecasting deficits, I find that expert forecasts or pooled forecasts always outperform the naïve model. From the pooled forecasts, the weighted forecast combination of all nine expert forecasters (WFC) is a combination that improves considerably over the single forecasts and nearly all other combined forecasts. This is because the WFC and Rbest are close in performance, as they also give more weight to the recently best performing experts. As private forecasters do not perform well, their weighted forecast combination (WFC) underperforms, and is beaten by any of the combined forecasts. I can further see that private forecaster N_5 has the worst performance: nearly all other combinations of forecasts beat this forecaster. This set of results shows that pooling may result in improvements in forecasting accuracy, and that those gains in accuracy are also statistically significant, in contrast to pooling time series models of the deficit (Favero and Marcellino, 2005).

The results in panel (b) show a rather different picture for the year ahead forecasts. None of the private or public forecasters beats the naive model - a result which is consistent with most of the findings in this area (Artis and Marcellino, 2001; Favero and Marcellino, 2005). Only the combination of forecasts is better than the naïve prediction (except the harmonic mean combination). Combining forecasts always beats the forecast of expert N_1, and surprisingly perhaps, also that of the EC. The literature has typically found that EC forecasts are better than other forecasts (Artis and Marcellino, 2001; Keereman, 1999), but I find that combining the forecasts for the year ahead deficit might produce gains in accuracy. Of all the combined forecasts, the weighted forecast combinations usually perform better than the private or public forecasters. The combination of all nine forecasters beats all forecasters, except for private forecaster N_4. However, there is no significant improvement

Table III.2. Diebold and Mariano test comparing single and combined forecasts to each combined forecast

a1) current year forecast

	N_1	N_2	N_3	N_4	N_5	MEF	OECD	EC	IMF	simple	harmonic	geometric	median	WFC	$disc_{95}$	$disc_{99}$	R_{best}	naive
simple	-0.06	-0.10	-0.11	0.01	-0.31	0.13	0.14	0.01	0.00									
	0.52	0.06	0.19	0.89	0.01	0.00	0.07	0.71	0.97									
harmonic	0.14	-0.09	-0.10	0.02	-0.32	0.14	0.13	0.01	-0.01	0.00								
	0.30	0.06	0.20	0.78	0.02	0.00	0.05	0.79	0.85	0.73								
geometric	0.08	-0.09	-0.10	0.01	-0.31	0.12	0.13	-0.01	-0.01	0.00	0.00							
	0.43	0.07	0.21	0.81	0.01	0.00	0.05	0.74	0.91	0.65	0.84							
median	0.10	-0.08	-0.08	0.04	-0.31	0.13	0.14	0.01	-0.01	-0.01	-0.01	-0.01						
	0.38	0.10	0.34	0.53	0.03	0.00	0.01	0.44	0.94	0.80	0.59	0.66						

a2) current year forecast

	N_1	N_2	N_3	N_4	N_5	MEF	OECD	EC	IMF	simple	harmonic	geometric	median	WFC	$disc_{95}$	$disc_{99}$	R_{best}	naive
WFC	-0.02	-0.42	-0.29	-0.02	-0.86	-0.16	-0.21	-0.06	-0.02	-0.32	-0.32	-0.32	-0.31					
	0,03	0,05	0,08	0,5	0.04	0.12	0.31	0,75	0,9	0.00	0.00	0.00	0.00					
$disc_{95}$	0.20	-0.17	-0.19	-0.05	-0.39	0.06	0.08	-0.05	-0.08	-0.05	-0.06	-0.05	-0.04	0.26				
	0.80	0.01	0.01	0.35	0.00	0.03	0.30	0.23	0.40	0.00	0.00	0.00	0.17	0.00				
$disc_{99}$	0.02	-0.17	-0.19	-0.04	-0.38	0.07	0.09	-0.04	-0.07	-0.05	-0.05	-0.05	-0.04	0.26	0.00			
	0.78	0.01	0.01	0.40	0.00	0.01	0.27	0.32	0.46	0.00	0.00	0.00	0.21	0.00	0.12			
R_{best}	-0.21	-0.42	-0.43	-0.28	-0.66	-0.19	-0.16	-0.31	-0.32	-0.29	-0.29	-0.29	-0.27	0.02	-0.24	-0.24		
	0.02	0.00	0.00	0.00	0.01	0.01	0.05	0.00	0.01	0.00	0.00	0.00	0.00	0.75	0.00	0.00		
naive	0.54	0.41	0.34	0.32	0.15	0.59	0.60	0.47	0.45	0.02	0.50	0.50	0.51	0.78	0.58	0.57	0.82	
	0.05	0.08	0.20	0.23	0.38	0.00	0.01	0.04	0.04	0.50	0.02	0.02	0.02	0.00	0.02	0.02	0.00	

b1) year ahead forecast

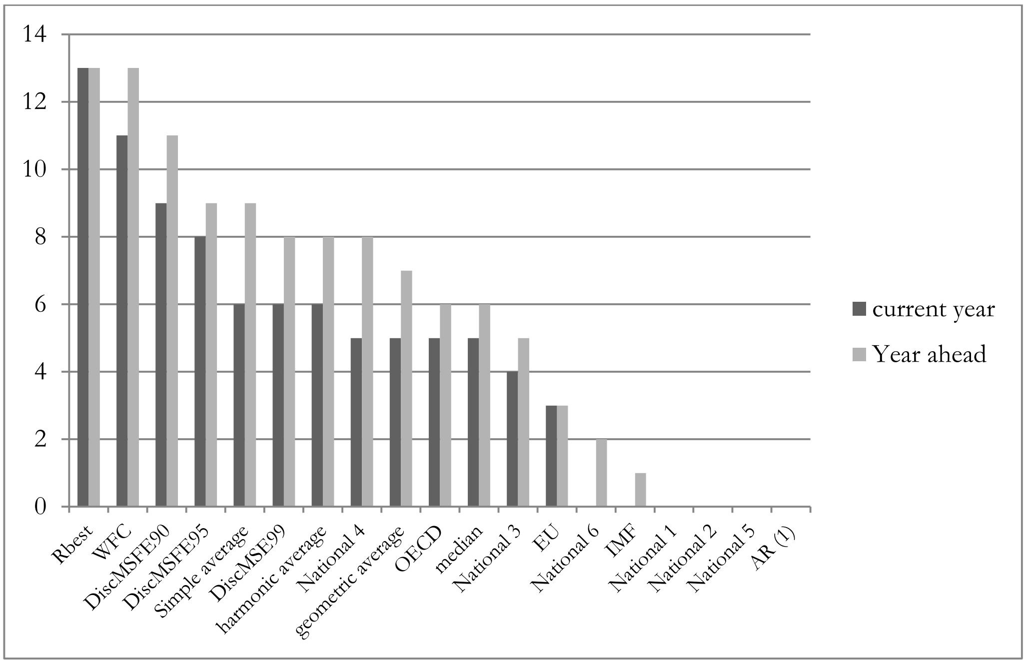
	N_1	N_2	N_3	N_4	N_5	MEF	OECD	EC	IMF	simple	harmonic	geometric	median	WFC	$disc_{95}$	$disc_{99}$	R_{best}	naive
simple	-0.27 0.01	-0.14 0.12	-0.01 0.84	0.07 0.21	-0.24 0.05	-0.01 0.85	-0.21 0.12	-0.29 0.01	-0.04 0.60									
harmonic	-0.19 0.01	-0.08 0.45	0.03 0.55	0.14 0.02	-.20 0.14	-0.12 0.44	-0.17 0.22	-0.26 0.03	-0.01 0.88	0.04 0.14								
geometric	-0.26 0.01	-0.11 0.27	0.01 0.92	0.10 0,07	-0.22 0.08	0.00 0.95	-0.20 0.15	-0.27 0.02	-0.02 0.71	0.01 0.16	-0.02 0.14							
median	-0.26 0.00	-0.12 0.26	0.03 0.14	0.08 0.04	-0.19 0.10	0.03 0.48	-0.16 0.31	-0.24 0.01	0.01 0.95	0.06 0.03	0.02 0.63	0.05 0.16						

b2) year ahead forecast

	N_1	N_2	N_3	N_4	N_5	MEF	OECD	EC	IMF	simple	harmonic	geometric	median	WFC	$disc_{95}$	$disc_{99}$	R_{best}	naive
WFC	-0.03 0,05	-0.86 0.03	-0.70 0.00	3.93 0.08	-.44 0.01	-6.18 0.00	-5.66 0.02	-0.46 0.02	-0.93 0.02	-0.06 0.62	-0.10 0.45	-0.08 0.55	-0.12 0.33					
$disc_{90}$	-0.26 0.00	-0.11 0.18	0.00 0.92	0.04 0.55	-0.25 0.06	-0.17 0.24	-0.23 0.08	-0.37 0.00	-0.06 0.40	-0.05 0.09	-0.08 0.05	-0.06 0.06	-0.08 0.00	0.04 0.77				
$disc_{95}$	-0.25 0.00	-0.11 0.17	0.00 0.95	0.04 0.51	-0.24 0.07	-0.16 0.27	-0.22 0.10	-0.36 0.00	-0.05 0.47	-0.04 0.12	-0.07 0.07	-0.05 0.08	-0.08 0.00	0.05 0.74	0.00 0.01			
$disc_{99}$	-0.24 0.00	-0.11 0.17	0.01 0.78	0.04 0.45	-0.23 0.08	-0.15 0.30	-0.21 0.12	-0.35 0.00	-0.04 0.54	-0.04 0.17	-0.07 0.09	-0.05 0.11	-0.07 0.00	0.05 0.73	0.01 0.06			
R_{best}	-0.32 0.00	-0.20 0.08	-0.06 0.20	-0.02 0.72	-0.31 0.05	-0.21 0.19	-0.27 0.07	-0.46 0.00	-0.15 0.16	-0.11 0.12	-0.14 0.09	-0.12 0.11	-0.15 0.01	0.00 0.98	-0.07 0.16	-0,07 0.13		
naive	0.22 0.50	0.40 0.11	0.42 0.07	0.33 0.30	0.20 0.09	0.28 0.23	0.23 0.24	0.14 0.55	0.39 0.10	0.47 0.03	0.43 0.06	0.46 0.03	0.41 0.04	0.53 0.01	0.54 0.03	0.54 0.04	0.54 0.04	

Note: for the coefficient of the regression (1) of the Weighted Forecast Combination (WFC) see Appendix A. Note: the first value is the statistic and under is the pvalue

Figure III.4. The rank of best performance of each model.



over other combination models. Few of the combination forecasts are actually significantly different from one another: only the discounted combination forecast (with δ=0.90) shows some difference in comparison to the other combinations.

To get a quick overview of the forecast that most often beats alternative ones, I plot the number of times that a forecast model's results are better than the others (not necessarily significantly so). The bars in Figure III.4 indicate this number for the current year and year ahead forecasts. In both cases, the Rbest and weighted forecast combinations do better than other forecasters at most times. Out of the 174 possible matches, Rbest comes out as the best 13 times for current year and year ahead forecasts. The WFC does so 11 times for the current year, and 13 times for the year ahead.

III.5. Stability in forecasts

III.5.1. Fluctuation test

A combination of forecasts aggregates information and reduces uncertainty by eliminating judgment errors on structural changes. The outcome is still based on the global performance of forecasters, however, and not on the change in performance of different competing forecasts over time. One of the reasons for the good performance of R_{best} or the weighted forecast combinations is that I select the best performing forecasters by their RMSE over the last couple of years. This time frame is sufficiently narrow to eliminate any forecasters that do not update their forecasts after structural changes. I observed in Figures III.1 and III.2 that forecasting performance on the Italian budget deficit changed over time. Up to 2001, most forecasters performed quite well, and projections were mostly aligned with actual budget outcomes. Afterwards, their performances diverged.

Since even expert forecasters are unable to anticipate all economic and political changes, forecasting models have to be adaptive. Finding an indicator that predicts well in one period is no guarantee that it will predict well in later periods. This explains the success of simple time series models in forecasting fiscal variables (Favero and Marcellino, 2005). More generally, expert forecasters using the same model are unlikely to outguess other experts at all points in time. Rather, the best forecasting model may change over time in ways that can be difficult to track on the basis of past forecasting performance (Timmermann, 2006).

Thus the Diebold and Mariano test (1995) is not the most best for carrying out forecast evaluations, as it favours models with short time horizons. If the predictive accuracy of a model relative to a competitor forecaster is very much connected to some specific period of time, I would like to test if its predictive accuracy changes over time. Giacomini and Rossi (2010) develop two tests that examine fluctuations in the relative predictive performance of two forecasting methods, A and B. Each method produces a sequence of out-of-sample forecasts based on a rolling window of m observations, which are used to construct the forecasting model at each point in time. At each point in time, I can then compute the difference in the accuracy loss of the two models as

$$\left\{\Delta L_t\left(\hat{\beta}_{t-h,R,}\hat{\theta}_{t-h,R}\right)\right\}_{t=R+h}^{T} \quad \text{(III.4)}$$

which depends on the realizations of the variable and on the in-sample estimates for each model re-estimated at each time $t=R+h,...T$ over a window of size R. The local relative loss for the two models is defined over centered rolling windows of size m as:

$$\frac{1}{m}\sum_{j=t-m/2}^{t+\frac{m}{2}-1}\Delta L_j\left(\hat{\beta}_{j-h,R,}\hat{\theta}_{j-h,R}\right) \quad \text{(III.5)}$$

In the first test, the sequence produced by repeated application of (III.5) allows us to evaluate the relative performance of both models at each point in time. The fluctuation test statistic is then defined as:

$$F_{t,m} = \hat{\theta}^2 m^{1/2}\sum_{j=t-m/2}^{t+\frac{m}{2}-1}\Delta L_j\left(\hat{\beta}_{j-h,R,}\hat{\theta}_{j-h,R}\right) \quad \text{(III.6)}$$

where the null hypothesis is that

$$H_0 = E\Delta L_t\left(\hat{\beta}_{t-h,R,}\hat{\theta}_{t-h,R}\right) = 0 \quad \text{(III.8)}$$

against a two-sided alternative whose performance is not similar. As with a structural break test, if the difference in performance exceeds a certain threshold in some time period, the null is rejected. Giacomini and Rossi (2010) derive the critical values (Table I, page 601) for testing the null hypothesis that the local relative MSFE equals zero at each point in time.

The fluctuation test does not specify an alternative hypothesis, so it may suffer from low power, as I do not know in which direction to look. I compute the MSFE differences over a rolling window of 10 years and test the null hypothesis that the MSFE is equal to zero for each combination model relative to the naive process. If the relative MSFE exceeds the critical value in some part of the sample, I reject the null hypothesis and I conclude that there are periods during the sample when one model outperforms the other. In addition, I run the one-time reversal test to check for changes in performance and the periods in which those occur. I test relative performance over time by comparing the nine public/private forecasts and the 11 combined forecasts.

Table III.3 reports the p-values of the null hypothesis in (III.7) for both the current year and year ahead forecasts. For the current year forecast, only the private forecasters – with the exception of N_1 – do not do significantly better than a naïve model in predicting the budget deficit. For all public forecasters, or a combination of forecasts, I can reject the null, so there is a significant gain associated with using these forecasts over a simple naive model. Given the full-sample results discussed before, it should not come as a surprise that the outcomes for the year ahead forecast of the budget deficit are quite different. In this case, none of the private or public forecasters beat the naive model. Only the forecasters N_3 and N_4, and the IMF, come close to beating the naive model, at 10 per cent. Instead, most of the combined forecasts do perform better than a naive model. For all of the combinations, I can reject the null at 10 per cent.

For the three types of weighted forecast combinations and the R_{best} forecast, this is even the case at 5 per cent. This result is interesting for two reasons. First, significant improvements in predictive power confirm previous findings that combination results in much more stable predictions. I show with this example that this result also holds in the case of updated predictions over time. Second, the typical finding in the literature on forecasting fiscal variables has been that simple time series models (or pooled versions of those) are among the few that perform better than a naïve model. I show that within the year, public expert forecasts or pooled versions produce substantial gains compared with individual forecasts and are robust to changes over time. But even more importantly, the pooling of expert year ahead forecasts nearly always results in improved performance that is also resilient robust in the face of structural changes.

Table III.3. P-values for Fluctuation Test with rolling windows of 10 years

Forecaster/Naive model	Current year	Year ahead
N_1	0.05	0.19
N_2	0.16	0.17
N_3	0.08	0.10
N_4	0.08	0.11
N_5	0.25	0.23
MEF	0.04	0.14
OECD	0.03	0.15
EC	0.05	0.21
IMF	0.05	0.11
simple	0.04	0.08
harmonic	0.04	0.10
geometric	0.04	0.08
median	0.04	0.06
WFC	0.02	0.02
disc90	0.03	0.04
disc95	0.03	0.05
disc99	0.03	0.05
R_{best}	0.02	0.03

III.6. Conclusions

Despite the growing importance of fiscal projections in the short-term to inform policy-makers, control fiscal monitoring, and manage expectations, practitioners seem to require a lot of judgment in making better fiscal projections. I show that exploiting the information from many different forecasters can still lead to substantial gains in predictive accuracy. Datasets that have become available in recent years, such as CEF, allow for the combining of forecasts in several ways. Applying eight different combination techniques to the current year and year ahead forecasts of the Italian budget deficit over the period from 1992 to 2012 results in substantial gains in forecast accuracy.

The results of the combination and the test show that the weighted combined forecast of the deficit ratio is superior to any single forecast. Deficits are hard to predict due to shifting economic conditions and political events. I test and compare predictive accuracy over time and although a weighted combined forecast is resilient to breaks, it does not significantly improve over a simple naive model. In particular, the combination models are more accurate than individual models for 65% in the year ahead and 54% in current year and, in any case, each combination model is better than a naïve one. Given the changes in forecast performance over time, no single model is to be preferred at any time, and a combination with the weighted forecast combination model provides the best performance compared by the other methods. Still, combining forecasts can result in substantial gains in predictive accuracy when compared against current standards.

III.7. References

Afonso, A., Schuknecht, L., & Tanzi, V. (2005) Public sector efficiency: an international comparison. *Public choice, 123(3-4),* 321-347.

Artis, M., & Marcellino, M. (1998). *Fiscal solvency and fiscal forecasting in Europe* (CEPR Discussion Paper No. 1836).

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal, 4,* 20-36.

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2013). *Forecasting fiscal time series using mixed frequency data* (European Central Bank Working Paper Series No. 1550).

Batchelor, R., (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics, 33*, 225 – 35.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*, 451-468.

Beetsma, R., & Giuliodori, M. (2010). Fiscal Adjustment to cyclical developments in the OECD: An empirical analysis based on real-time data. *Oxford Economic Papers 62*(3), 419-441.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*, 559-583.

Clemen, R.T., & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics, 4,* 39-46.

Clements, M. P., & Hendry, D. F. (2003). Economic forecasting: some lessons from recent research. *Economic Modelling, 20*(2), 301-329.

Cooper, J.P., & Nelson, CR. (1975). The ex-ante prediction performance of the St. Louis and FRB-MIT-PENN econometric models, and some results on composite predictions. *Journal of Money, Credit and Banking,7,* pp. 1–32.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253-263.

Dovern, J., & Weisser, J. (2011). Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: an empirical comparison for the G7. *International Journal of Forecasting, 27*, 452 –65.

Favero, C., & Marcellino, M. (2005). *Modelling and forecasting fiscal variables for the Euro Area* (Bocconi University Working Paper No. 298).

Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics, 24,* 435- 68.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25,* 595–620.

Granger, C. W. J., & Newbold, P. (1973). Some comments on the evaluation of forecasts. *Applied Economics, 5*, 35-47.

Jonung, L., & Larch, M. (2006a). Fiscal policy in the EU: are official output forecasts biased? *Economic Policy, July*, 491-534.

Jonung, L., Larch, M. (2006b). Improving fiscal policy in the EU: the case for independent forecasts. *Economic Policy, CEPR; CES; MSH, 21*(47), 491-534.

Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *American Economic Review, 80,* 714–735.

Keereman, F. (1999). The track record of the Commission forecasts (European Commission, Directorate-General for Economic and Financial Affairs Economic Paper No. 137).

Leal, T., Pérez, J. J., Tujula, M., & Vidal, J. P. (2008). Fiscal forecasting: Lessons from the literature and challenges. *Fiscal Studies, 29*, 347–386

Lupoletti, V. M., & Webb, R. H. (1986). Defining and improving the accuracy of macroeconomic forecasts: Contributions from a VAR model. *Journal of Business 5*(9), 263-285.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting, 1*, 111-153.

Makridakis, S., Winkler, R.L., (1983). Averages of forecasts: Some empirical results. *Management Science,* 987-996.

Makridakis, S., (1989). Why combining works? *International Journal of Forecasting, 5*(4), 601-603.

Marcellino, M. (2002). *Forecast pooling for short time series of macroeconomic variables* (CEPR Working paper No. 3313).

Marcellino, M. (2004). Forecast Pooling for European Macroeconomic Variables. *Oxford Bulletin of Economics and Statistics, 66*(1), 91-112.

Nelson, C.R. (1972). The prediction performance of the F.R.B.-M.I.T.-PENN model of the U.S. economy. *American Economic Review, 62*, 902-917.

Onorante, L., Pedregal, D., Pérez, J. J., & Signorini, S. (2010). The usefulness of infra-annual government cash budgetary data for fiscal forecasting in the euro area. *Journal of Policy Modeling, 32*(1), 98-119.

Ottaviani M., & Sorensen, P. N. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, *81*, 441–466.

Pedregal, D., & Pérez, J. J. (2010). Should quarterly government finance statistics be used for fiscal surveillance in Europe? *International Journal of Forecasting, 26(*4), 794-807

Pesaran, M., Yongcheol, S., & Smith, R. (2004). *Pooled mean group estimation of dynamic heterogeneous panels* (Edinburgh School of Economics Discussion Paper No. 16).

Stock J. H., and Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Perspectives*, *41*, 788–829.

Stock J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*, 405-430.

Theil, H. (1958). Chapter 2, in *Economic forecasts and policy*. Amsterdam: North Holland.

Timmermann, A. (2006). Forecast Combinations. In Elliot, G., Granger, C. & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (136-196).

Trueman, B. (1994). Analyst forecasts and herding behaviour. *Review of Financial Studies, 7*(1), 97–124.

Winkler, R. L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, 146*, 150-157.

Zamowitz, V. (1967). *An appraisal of short-term economic forecasts* (National Bureau of Economic Research New York, 123-125).

Annex III.1. Computation of weighted forecast combinations

The weighted forecast combination is based on an OLS pooled regression of the realised deficit on the various forecasts over the full sample. The weights are nothing else than the coefficients, and the combined forecast fits the model.

I compute the forecast combinations, based on a regression WFC of all nine forecasters. The following table reports the regression results for the current year and year ahead forecasts. I use all available forecasts from 1992-2012, but drop missing values.

Table A.III.1 Weights for weighted forecast combinations.

	database Spring		database Autumn	
	Forecast Current Year	Forecast Year Ahead	Forecast Current Year	Forecast Year Ahead
C	0.59 (1.72)	0.78 (1.64)	0,65 (1.68)	0,42 (1,52)
N2	-0.61 (-1.44)	0.42 (0.98)	-0.70 (-1.35)	0.45 (0.96)
N5	-0.54 (1.68)	0.15 (0.34)	-0,15 (1,55)	0,26 (0,32)
MEF	0.81 (1.6)	1.22 (1.00)	0,12 (1,4)	-0,23 (1,12)
OECD	0.56 (0.6)	-0.87 (-0.71)	0,24 (0,8)	0,49 (0,65)
EC	0.04 (0.05)	0.02 (0.06)	0,2 (0,02)	0,19 (0,05)
IMF	0.64 (1.24)	-0.04 (-0.07)	0,55 (1,15)	0,18 (0,08)
naive	-	-	-	
R2	0.94	0.87	0,9	0,88
F	36.42	15.54	45,2	17,25

Note: the numbers in parenthesis are the t-statistics. “Spring” refers to the months May/June and “Autumn” refers to the months October/December as a robustness check.

Annex III.2. Calculation of the forecasted budget balance (as a ratio of GDP)

CEF only provides forecasts for the total deficit in nominal values (local currency). Hence, I follow Heppke-Falk and Hüfner (2004) and Poplawski-Ribeiro and Rülke (2011) in constructing a forecast measure of the deficit to GDP ratio (percentage of GDP). To do so, I cannot simply scale the nominal value deficit forecast by the GDP forecast, since the CEF surveys for growth rates only, and not for the GDP in nominal value.

I construct a measure of the expected nominal year ahead GDP forecast of forecaster i at month m and year t as follows. In the first step, I take a real-time measure of real GDP levels for a particular year t. I use the real-time forecast of the same-year real GDP (in levels) coming from the most recent IMF World Economic Outlook (WEO) vintage available at any particular month m of year t. The IMF WEOs are published either in April or October, hence from May to October I use the April issue, and in the other months I use the October issue.

The second step is to compute the nominal value of the year ahead GDP forecast. I multiply the real-time (WEO) measure of same-year real GDP (in levels), $E_{WEO,t}\,[y_t]$, by the year ahead market (Consensus) forecasts for GDP *growth*, $E_{i,t,m}[\Delta y_{t+1}]$, and inflation, $E_{i,t,m}[\pi_{t+1}]$, for each forecaster i at a particular month m of year t. The expected year ahead nominal GDP value for each country is then

$$E_{i,t,m}[y_{t+1}] = E_{WEO,t}\,[y_t] \text{ x } (1 + E_{i,t,m}[\Delta y_{t+1}] + E_{i,t,m}[\pi_{t+1}]). \qquad \text{(A.III.2.1)}$$

The year ahead expected budget balance for each country is then:

$$E_{i,t,m}[b_{t+1}] = \frac{E_{i,t,m}[b^{\,nom}{}_{t+1}]}{E_{i,t,m}[y_{t+1}]} \qquad \text{(A.III.2.2)}$$

where $E_{i,t,m}[b^{\,nom}{}_{t+1}]$ is the (CEF) forecast of the nominal budget balance by forecaster i in month m of year t for one year ahead $t+1$.

Chapter IV.

Nowcasting public finance in Italy[13]

IV.1. From forecasting to nowcasting

Nowcasting is a technique applied to high frequency data with the aim of forecasting the public deficit in Italy. To apply this methodology, I analyse a series of monthly public finance indicators such as the Economic confidence index and Google trends. The purpose of this work is to make available efficient and accurate forecasts of fiscal deficit trends in those months in which official forecasts are not published. I compute a real-time deficit value from different parsimonious time series models. The idea is to provide an econometric tool to support the actual planning and the monitoring of infra-annual forecasts of budget variables. Building parsimonious models with high frequency data could provide information about different aspects of public finance (Bianchi et al., 2010), such as:

- planning: a viewpoint of infra-annual payments and cash flows;
- monitoring and financial management: update of short term monthly forecasts based on monitoring during the financial year;
- simulation: intervention strategy on variations of budget assignments,
- public savings: forecasting the unexpended balance at the end of the financial year.

The literature covered in Chapter I is in line with the recent Italian approach of nowcasting. The techniques of the monthly series analysis in Italy are taken by Pozzuoli and Ercoli (2008). They study the time series from 2002-2007 with the aim of forecasting the short term financial requirement trend. In particular, they use a monthly time series of budget variables and analyse the statistical properties of the data; then they eliminate the seasonality components to make

[13] I presented this Chapter at PhD seminars at the Universitat de Barcelona in December 2014 and at Seminar sessions at the Università Tor Vergata of Rome (Italy) in February 2015 and at the ISF 35th International Symposium on Forecasting in Riverside (California) in June 2015.

the variables stationary. They run models on the financial requirements with a univariate SARIMA equation and provide good results from one to six forecast years ahead. They make the same analysis at the aggregate level in terms of the difference between expenditures and revenues, and consider them separately. The results show that the forecasts' balance improves on the difference between the forecast expenditures and revenues. In particular, different approaches are used. The direct method is to nowcast the aggregate variables using only their own past specifying an autoregression (AR, ARIMA) using limited informations of low frequency Mitchell et al., (2005), Ferrara et al. (2010), and Bànbura et al. (2010, 2012). Other direct methods are the Linear regression model and the Vector Autoregression Model (VAR), used by Blanchard and Perotti (2002) and Perotti (2002) to forecast macroeconomic variables.

Nowcating can also be based on combining forecasts of disaggregated components (Hendry and Hubrich, 2011) depending on whether the nowcasted variable is known or not. Nowcasting could be difficult when there are missing values or the variables are published with different frequencies. In this case, Clements and Hendry (2003) reviewed the methodology and introduced the "in-filling" of missing disaggregates based on exponentially weighted moving average or an autoregressive-integrated moving average. The problems of the different frequencies are solved by Ghysels et al. (2004, 2007, 2012) who proposed the mixed-data sampling approach (MIDAS), which links a low-frequency variable with selected predictors at higher frequency using a parsimonious restricted lag polynomial. Another approach is due to a large number of explanatory variables. Stock and Watson (2002) solve the problems through the combinations of fiscal forecasts and also summarize the dynamics of the monthly indicators using a small set of common factors, called "factor models", and apply those as explanatory variables. Boivin and Ng (2005) provide a summary of factor models for single frequency data. These two latter methods are parsimonious but restrictive ways of combining mixed frequency data and ragged edges, but the results are difficult to interpret. For this and other reasons, another approach uses the "Bridge equations" (Perez, 2007). This approach is based on equations that construct a direct bridge between aggregate measures and a set of explanatory variables of different frequencies. Banbura et al. (2010) propose modelling the monthly data as a parametric dynamic factor model cast in a state space representation. After obtaining the state space representation they use the Kalman filter techniques to perform the projections as they automatically adapt to changing data availability.

In the literature, different works perform similar exercises with different fiscal variables, Ghysels and Ozkan (2012) use ADL-MIDAS regression models to obtain forecasts for U.S. federal government expenditures, revenues, and deficits both at quarterly and annual frequencies, and find good results on autoregressive models and ADL, but not in monthly frequency. Other variables that are nowcasted in the literature include GDP in Irish (D'Agostino, et al., 2011), GDP growth, inflation (Banerjee et al., 2006), and unemployment in the US economy using the DSGE model and BVAR, which outperform the RW (Smets et al., 2013). Pérez (2007) provides a number of fiscal indicators based on monthly and quarterly public accounts, which could anticipate the annual fiscal deficits. Other international approaches, for example Moulin et al. (2004), have focused on the forecast public balance based on autoregressive models in France. Their results improve the forecasts provided by the Government. Further studies focused on the Euro Zone and Spain (Pedragal and Perez (2010), Leal et. al (2009), Onorante et al. (2010) use models at mixed frequencies to monitor public deficit forecasts. Castle et al. (2013) provide a deep methodological overview.

The empirical results of the present Chapter contribute to the recent literature in different aspects, for example using econometric models for nowcasting public finance and to provide a complete database to this aim. Indeed, these data are high frequency variables on public finance, economic indices, and Internet searches. The results obtained have a dual utility methodological and policy; they are useful methodologically, in the sense that applied nowcasting techniques can be applied to the public deficit, and also this is also important for economic policy. Econometrically, the use of no – parametric Wilcoxon test shows that all of the models capture the deficit trend and perform better than a naïve model. In particular the Linear Regression Models are more accurate at forecasting the increases or decreases in the public deficit. The policy implications of this work could be to provide useful insights in terms of assessing the impact of financial deficit on the budget policy.

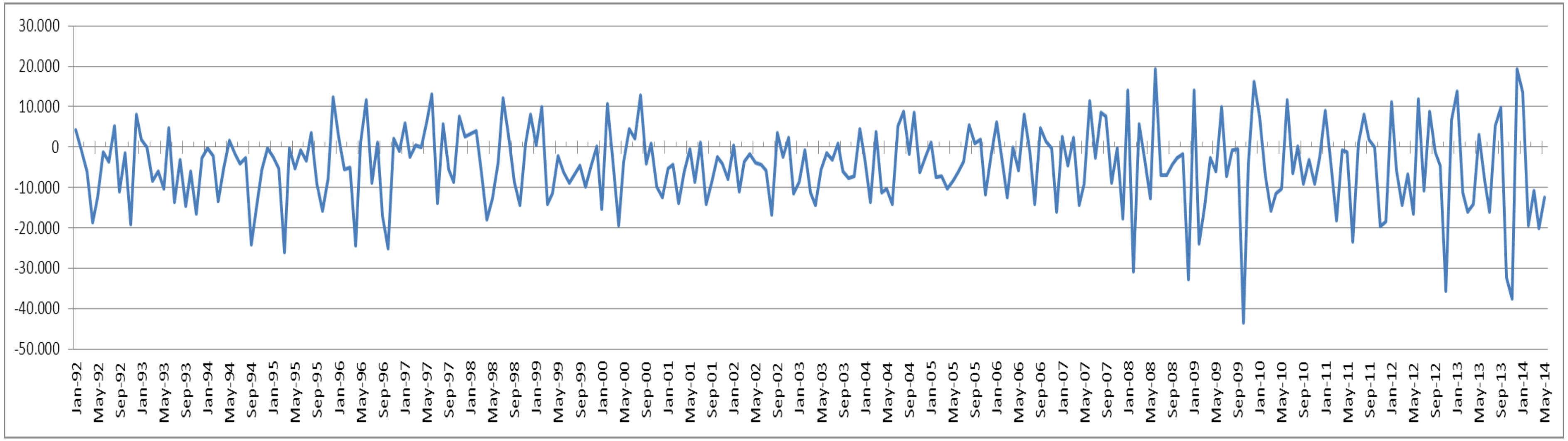
The Chapter is organized as follows. In Section IV.2, I describe the different datasets, while in Section IV.3, I present the different econometric models and the way I compute the nowcasts. In Section IV.4, I present the main results of the non-parametric test. Section IV.5 concludes the Chapter.

IV.2 Databases

The deficit that I aim to forecast is a monthly series from 1992:01 - 2014:05 in Italy released by the Bank of Italy and includes the balance of change in central government liabilities, excluding those that are assets of general government entities, and the change in the Treasury's liquid balances. State transfers to other general government entities contribute to the formation of the central government's borrowing requirement. The series includes liabilities connected with loans granted to countries belonging to the Economic and Monetary Union via the European Financial Stability Facility. The public finance data relating to the central government are timelier than those relating to the performance of the economy in general and are the best proxy of the government's annual general lending. For example, in the case of expenses, Giovannini (1991) had found that the introduction of general government variables in equations improved his models but made their forecasting performance worse due to the variables delay compared to central government finance variables (Bianchi et al., 2010). The value of the deficit that I analyse in this chapter is expressed in millions of euros. Figure IV.1 shows the series. I can see the high variability of the deficit[14] over the last two decades. Indeed, a strong drop of the variable is registered during the crisis from 2008 onwards. For this reason, I seasonally adjust the deficit series and check for AC and PAC to verify the order of integration (see Annex IV.3.).

[14] Central Government Lending is more prompt in monthly series compared with the other economic variables.

Figure IV.1. Monthly deficit in Italy 1992:01-2014:05



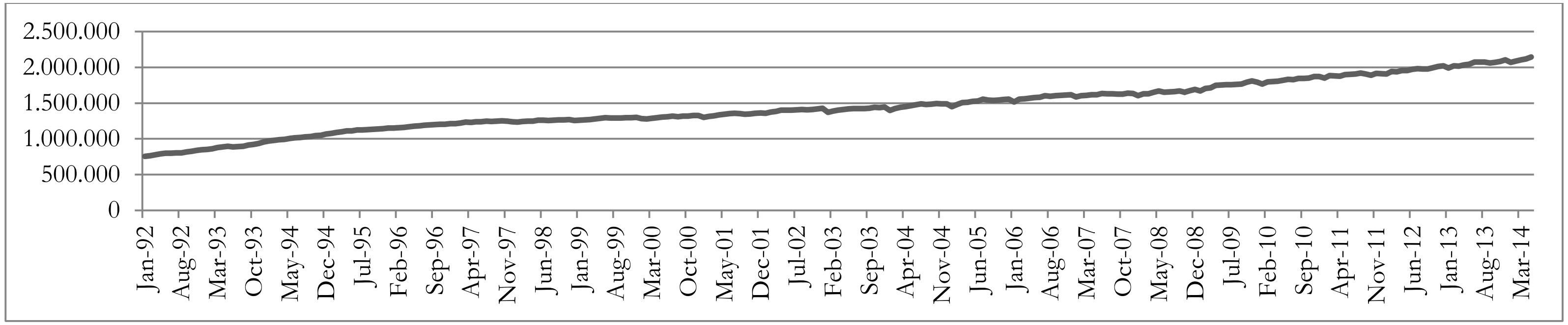
Note: the variable is named "cash balance" and it is a proxy of deficit in terms of liquidity. Indeed, deficit as economic budget evaluation it's not available monthly.

In order to model this deficit series, I use monthly data from the same deficit time range of (1992:01-2014:05) for debt, inflation, the production index, and the short term interest rate from the Bank of Italy dataset. These are the variables included in the works of Favero and Marcellino (2005) and Ghysels and Ozkan (2012) shown in figure VI.2. The graphs show, in general, a worsening of economics in real and monetary terms. In particular, inflation tends to decrease with two strong downfalls in July 1997 and 1998 carrying on a diminishing of monetary power during this time. The index of production tends to increase with a critical down at the beginning of 2009 showing the consequences of the economic crises on industry. Debt tends to increase constantly during the entire sample.

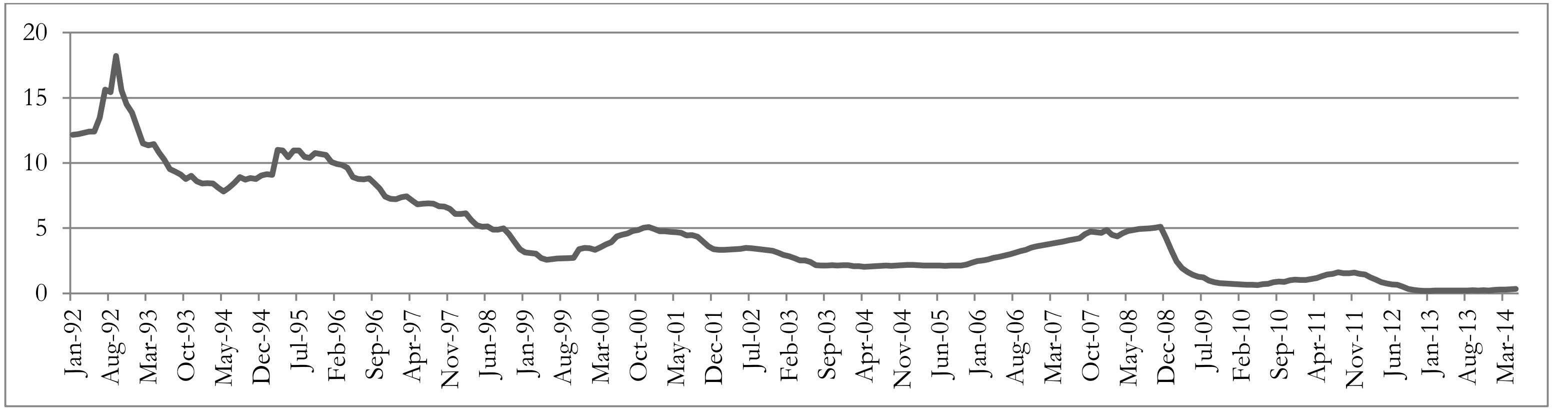
To complement the analysis, I further select a series that reflects the current stance of the Italian economy. This consider the composite business confidence climate index (IESI, ISTAT, Economic Sentiment Indicator), obtained by summarizing the confidence climates of manufacturing, construction, market services, and retail trade. This series is from the Italian National Institute of Statistics (ISTAT). The relative Figure shows that this series has very interesting behaviour: in fact, people's trust in the economy changed after 2008. Until then, the climax index showed low levels of trust but after the 2008 crisis, the people's emotional response about the economy tended to increase. These two groups of variables are considered in the sample 1992:1-2014:4.

Figure IV.2. Macro variables

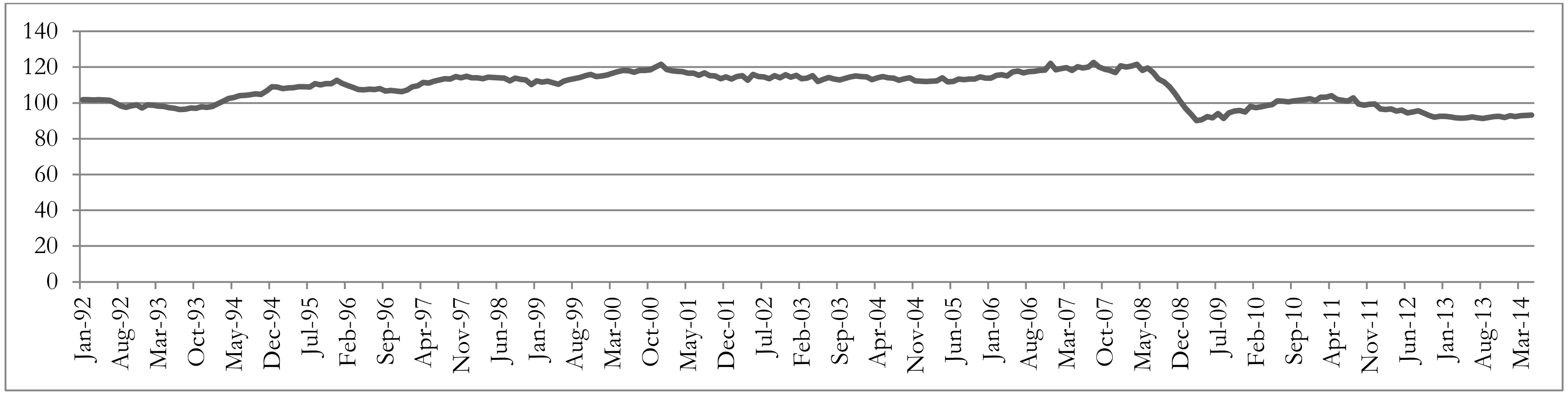
Debt



Short term interest rate

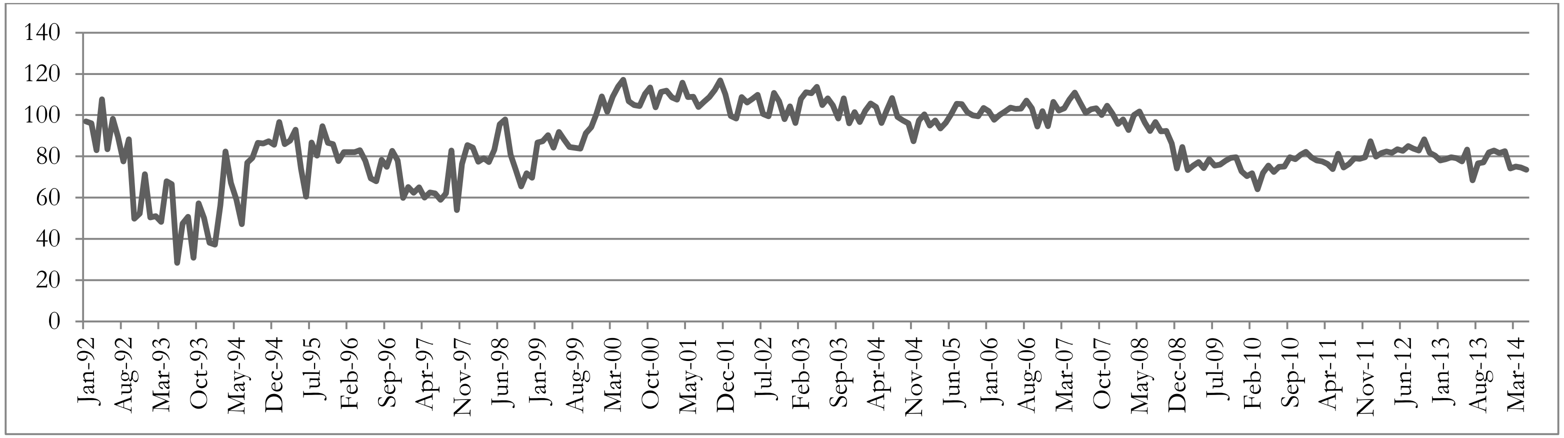


Index of Industrial production



Inflation

Economics confidence index



I also experiment with a group of online search terms. I have analysed the Internet source in order to obtain a series of data that include leading indicators that are fast and reliable. For this purpose I used the Google's "Insight for Search" web application, where by entering keywords in the search criteria, the Google engine allowed me to capture information for a particular query of a country or a region. These data represent the Google trends, it means the number of times a keyword is searched in the search box; the returning output measures the volume of interest by the operators of that particular topic. Additionally, search queries are useful for finding news, issues of social concern, and up to date information. Google provides yearly, monthly, weekly, and even real-time data. The first data available are from 2004. Google normalizes the data to a value between 0 and 100, where the first is associated with no search of the topic and the latter corresponds to high frequent a word is searched.

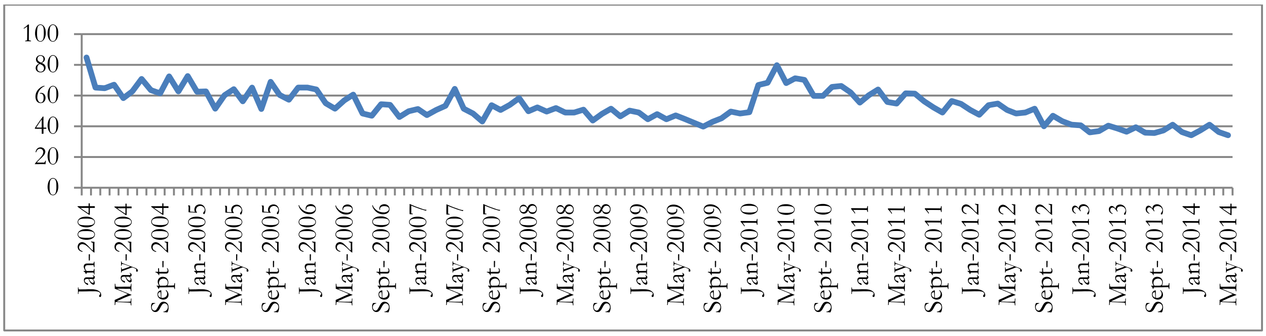
I use Google's index as the best leading indicator of web searching, following the work of D'Amuri and Marcucci (2009), Askitas and Zimmerman (2015), who use these index to predict unemployment and private consumption through keyword searches. I use the same approach to predict deficits. As they do in their work, I select the most popular keywords by looking at how often they are used. I look in particular at the following keywords that could be most relevant for the budget: "Tesoro" (Treasury), "Bilancio"(Budget), "Finanziaria" (Finance Low) and "Riforma"(Reform).

The series are obtained from Google Trends, an online tool from Google that measures the number of Internet searches made for these words with the Google search engine each month. In particular, I found that the word "Finanziaria" was the one that was used most often among different finance search keywords.

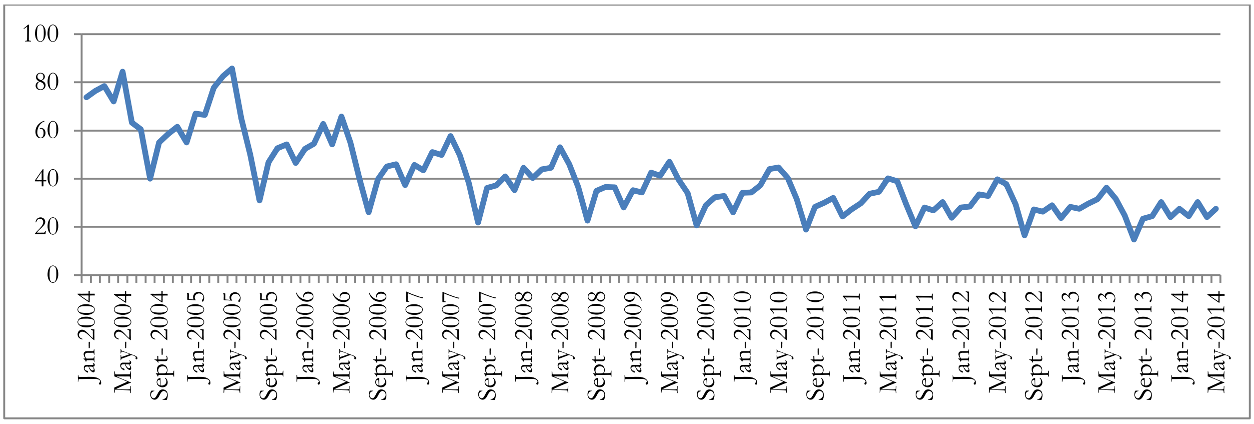
For my analysis, I downloaded weekly data and converted it into monthly data to allow comparison with the other two datasets (financial and climax variable). I calculated the average of four continuous weeks, with the fourth week containing the last day of the month.

Figures IV.3. Google trends

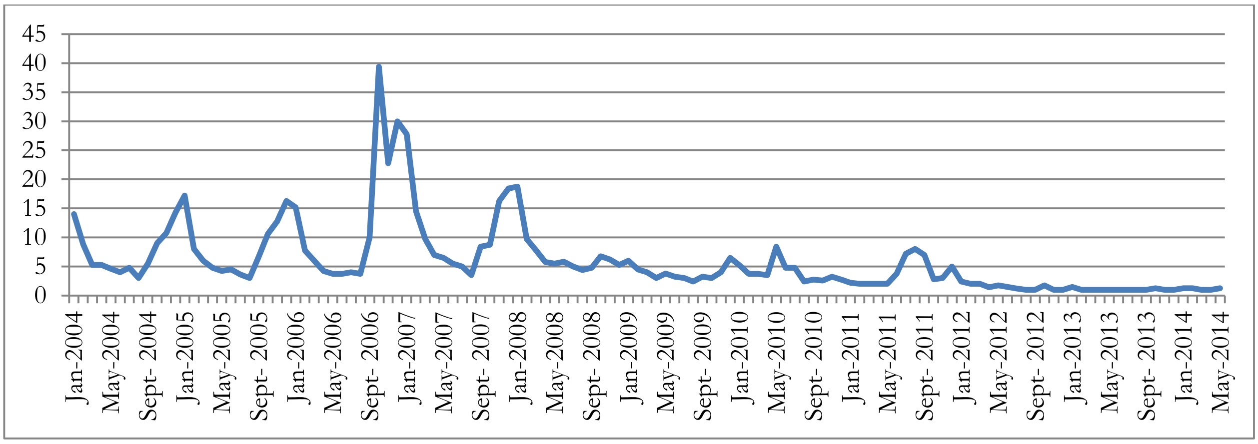
“Tesoro”



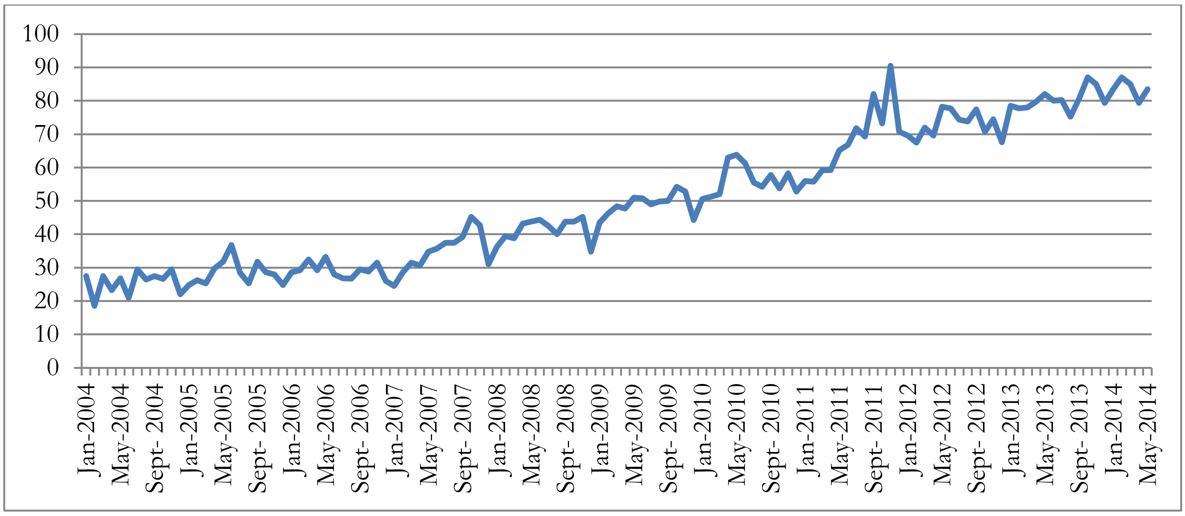
“Bilancio”



“Finanziaria”



“Debito”



“Riforma”

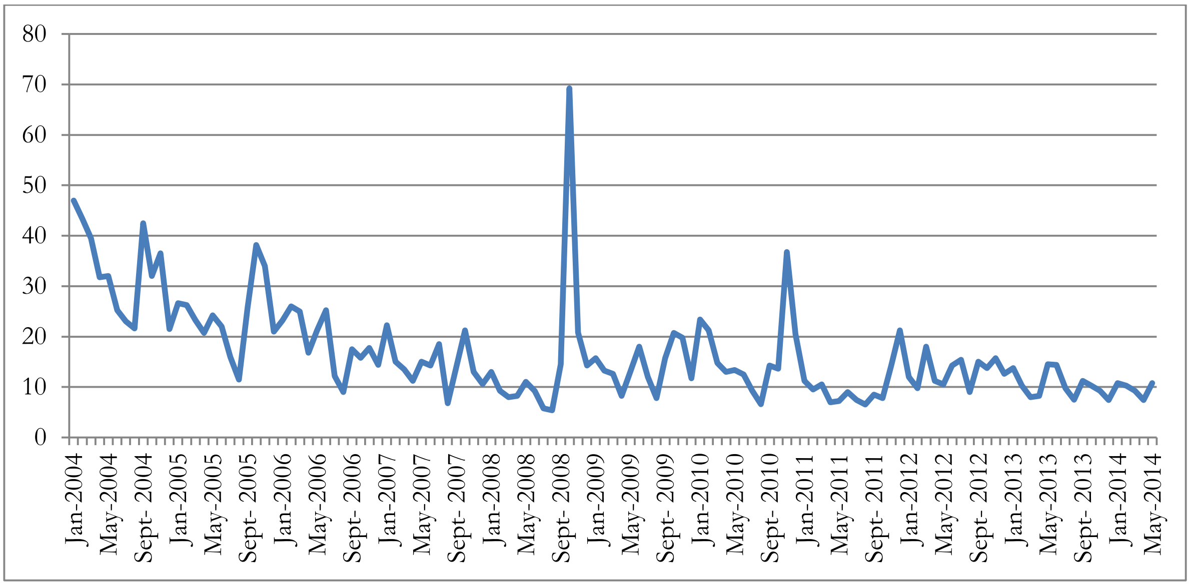


Table IV.1. Definition variables, sources, and time of publication

Variables	Definition	Publication	Institution
Deficit	Cash balance of the state budget related to central current and capital expenditure. It excludes transactions relating to recognition and the repayment of loans.	Updated in the middle of each month	Bank of Italy
Debt	The sum of general government financial liabilities excluding those that are assets of general government entities.	Updated in the middle of each month	Bank of Italy
Index of Production	Index measuring the variation of the production volume by industry.	Updated in the middle of every three months	ISTAT
Short term interest rate	Short term rates are usually either the three month interbank rate attached to loans given and taken among banks for any excess or shortage of liquidity over several months, or the rate associated with Treasury bills, Certificates of Deposit, or comparable instruments, each having a three month maturity	Updated in the middle of every two months	OECD
Inflation	Index with the entire population included. It includes the whole of the goods and the services bought by families with a market price. It is the tool to measure inflation.	Updated at the beginning of each month	ISTAT
Climax index	It is the index of the results investigation. Its goal is to evaluate the optimism/pessimism of the Italian family.	Updated at the beginning of each month	ISTAT
Google Trends	Google searches to compute how many times a term is entered, relative to the total number of searches made on Google over the same time frame.	Updated daily	Goolge Inc.

I generated monthly series of Google search query trends that show the volume of searches on themes related to deficit. As mentioned earlier, this was the only available data in the sample 2004: 1-2014: 4.

In Figure IV.3, all the Google indices show a peak in 2008, but they have different trends. Google "Tesoro" decreases until 2008, then it has a peak and tends to decrease again; Google "Bilancio" decreasing constantly; Google "Finanziaria" and Google "Riforma" have two peaks at the end of 2007 and in 2008 and then became constant; and Google "Debito" increases in all the samples.

In table IV.1, I summarize the main variables that I use in my analysis, the sources, and the timing of publications. The deficit and debt are updated in the middle of each month and published by the Bank of Italy; the Index of production is updated in the middle of each three month period and published by Istat; the short term interest rate is updated in the middle of each two month period and published by the OECD; and the inflation is updated at the beginning of each month and produced by Istat as as the Economic Confidence Index. The Google trends are updated four times a week and published online.

Finally, I use the dummy variables to control the pattern of the monthly deficit. This choice, in general, is justified by the seasonality of the actual deficit, and also by the potential impact of changes in government the government changes, which I can summarize in Table IV.2. The impact of political cycles on forecasting accuracy has been formally tested by Beetsma et. al. (2012), who find that political factors, specifically during elections, influence the optimism of forecasts. Also, Pina and Venes (2011) consider the importance of elections on overpredictions while Merola and Perez (2013) find that electoral cycles have a significant influence on governmental and international fiscal forecasts. One potential explanation of this result, also found here, is that forecasters perhaps try to minimize the impacts of potentially different economic policies after elections on their forecast. In table IV.1, I show the months of the year in which changes in government occur. Looking at the deficit series in Figure IV.2, the electoral cycle seems to predominate around 2001, when Italy came to be part of Euro zone, when there was the Berlusconi II Government, and during 2006-2008, when I observe a trend of higher volatility of the series, perhaps due to an incorrect assessment of the impact of the Great Recession on the Government of Berlusconi III and Prodi II. As Cimadomo (2012) argues, during the milder downturns (and upturns) that characterized the economic cycle in advanced economies before the 2008–2009 crisis, discretionary fiscal measures that were forecasted ex-ante with a counter-cyclical aim might have been pro-cyclical from an ex-post perspective.

Table IV.2. Changes in Government Legislature from 1992 to 2014

Government	**Legislature**
Amato *	28.06.1992 - 28.04.1993
Ciampi *	28.04.1993 - 10.05.1994
Berlusconi I **	10.05.1994 - 17.01.1995
Dini *	17.01.1995 - 17.05.1996
Prodi *	17.05.1996 - 21.10.1998
D'Alema *	21.10.1998 - 22.12.1999
D'Alema *	22.12.1999 - 25.04.2000
Amato *	25.04.2000 - 11.06.2001
Berlusconi II **	11.06. 2001-23.04.2005
Berlusconi III **	23.04.2005-17.05.2006
Prodi II *	17.05.2006-6.05.2008
Berlusconi IV **	8.05.2008 -16.11.2011
Monti tec	16.11.2011-27.04.2013
Letta *	28.04.2013 – 24.02.2014
Renzi *	from 24.02.2014

Source: my elaboration. Note: * left-wing parties, ** right-wing parties and tec is technical Government

IV.3. The models

IV.*3.1. Econometric methods to estimate the monthly fiscal deficit*

I set up models of time series to forecast the monthly deficit, following Favero and Marcellino (2005) and Ghysels and Ozkan (2012). These are different types of time series models (AR, ARMA, ARIMAX), Linear regression models (LRM), and Vector autoregressive model (VAR).

I construct an AR(p) model of the deficit d_t, with an order p, which can be written as:

$$X_t = c + \sum_{i=1}^{p} \varphi X_{t-1} + \varepsilon_t \qquad \text{(IV.1)}$$

Where $\varphi_{1,}, \varphi_{2,} \varphi_{3,} \dots \varphi_{n,}$ are parameters, c is a constant, and the random variable ε_t is white noise.

I experiment with an AR(12), AR(3), and AR(4) model, including a correction for the seasonality of the series.

To extend the model to an *ARMA Autoregressive moving average model,* the notation ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms. This model contains the AR(p) and MA(q) models,

$$X_t = c + \varepsilon_t \sum_{i=1}^{p} \varphi X_{t-1} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-1} \qquad \text{(IV.2)}$$

I specify a SARIMA(1,0,1,12) and SARIMA(1,0,0,12) with a constant and seasonality included.

When I add different indicators to the model, I estimate an ARMAX model. The notation ARMAX(p, q, b) refers to the model with p autoregressive terms, q moving average terms, and b exogenous inputs terms.

This model contains the AR(p) and MA(q) models and a linear combination of the last b terms of a known and external time series s_t(dummy). It is given by:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1} + \sum_{i=0}^{b} \eta_i s_{t-1} \quad \text{(IV.3)}$$

Where $\eta_1, \eta_2, \ldots, \eta_n$ are the parameters of the exogenous input s_t. In the simplest case, I simply do an ARMAX(1,1,1) with seasonality included.

Single equation models often perform well, thanks to their resilience to structural shocks. A drawback is that the model cannot account for a lot of other influences of other variables. For that reason, Favero and Marcellino (2005) use a VAR model. A VAR model describes the evolution of a set of k endogenous variables over the same sample period (t = 1, ., T) as a linear function of only their past values. The variables are collected in a k × 1 vector y_it, which has the T element, y_i,t, the time t observation of the i variable. In my case, I include the deficit series, as well as the index of industrial production:

$$X_t = c + A_t Y_t \quad \text{(IV.4)}$$

where X_t is the dependent variable at time t, c is the constant and A_t is the vector, and Y_t is the independent variable.

A) Regression models using Ordinary Least Square:

$$X_t = \alpha + \beta_t Y_t + \varepsilon \quad \text{(IV.5)}$$

Where X_t is the dependent variable deficit at time t. The $Y_{t,}$ collect the parameters of the exogenous strictly economic inputs as the index of industrial production, the short-term interest rate, the inflation, and the climax index with the list of Google trends. I can summarize all the models in table IV.3.

Table IV.3. The models

Model 1	AR(12)
Model 2	AR(3)
Model 3	AR(4)
Model 4	Var (with LOG INDEX OF INDUSTRIAL PRODUCTION)
Model 5	SARIMA (1,0,1,12)
Model 6	SARIMA (1,0,0,12) + COSTANT
Model 7	SARIMA (1,0,1,12) +COSTANT
Model 8	SARIMA (1,0,1,12)
Model 9	SARIMA (1,0,0,12) + ARIMAX WITH LOG INDEX OF INDUSTRIAL PRODUCTION
Model 10	REGRESSION (LOG INDEX OF INDUSTRIAL PRODUCTION, DEBT, INFLATION, INTEREST)
Model 11	REGRESSION (LOG INDEX OF INDUSTRIAL PRODUCTION, ECONOMIC CONFIDENCE INDEX)
Model 12	REGRESSION (LOG INDEX OF INDUSTRIAL PRODUCTION, ECONOMIC CONFIDENCE INDEX, GOOGLE'S TRENDS)
Model 13	REGRESSION (LOG INDEX OF INDUSTRIAL PRODUCTION, DEBT, INFLATION, INTEREST, GOOGLE'S TRENDS)

IV.*3.2. Nowcasting the deficit*

In this section I will present the methodology of nowcasting. The models are based on the full sample 1992:01-2013:12. Then, each model is re-estimated over the nowcast period 2014:01-2014:05. The estimation produced four-month nowcasts, both static and dynamic. I also estimate the recursive model rolling forecasting exercise as static and dynamic. In particular, I perform models of monthly deficit in the last year of the sample with a window of one month. I compute the models starting from the estimate origin from 1992:01 to 2013:05. Then I forecast the model over rolling windows from 2013:06 to 2013:10. Then I maintain the same forecast origin from 1992:01 and I move the sample to the next one month 2013:6, then nowcast from 2013:07 to 2013:11. I do the same computation covering one year of the sample.

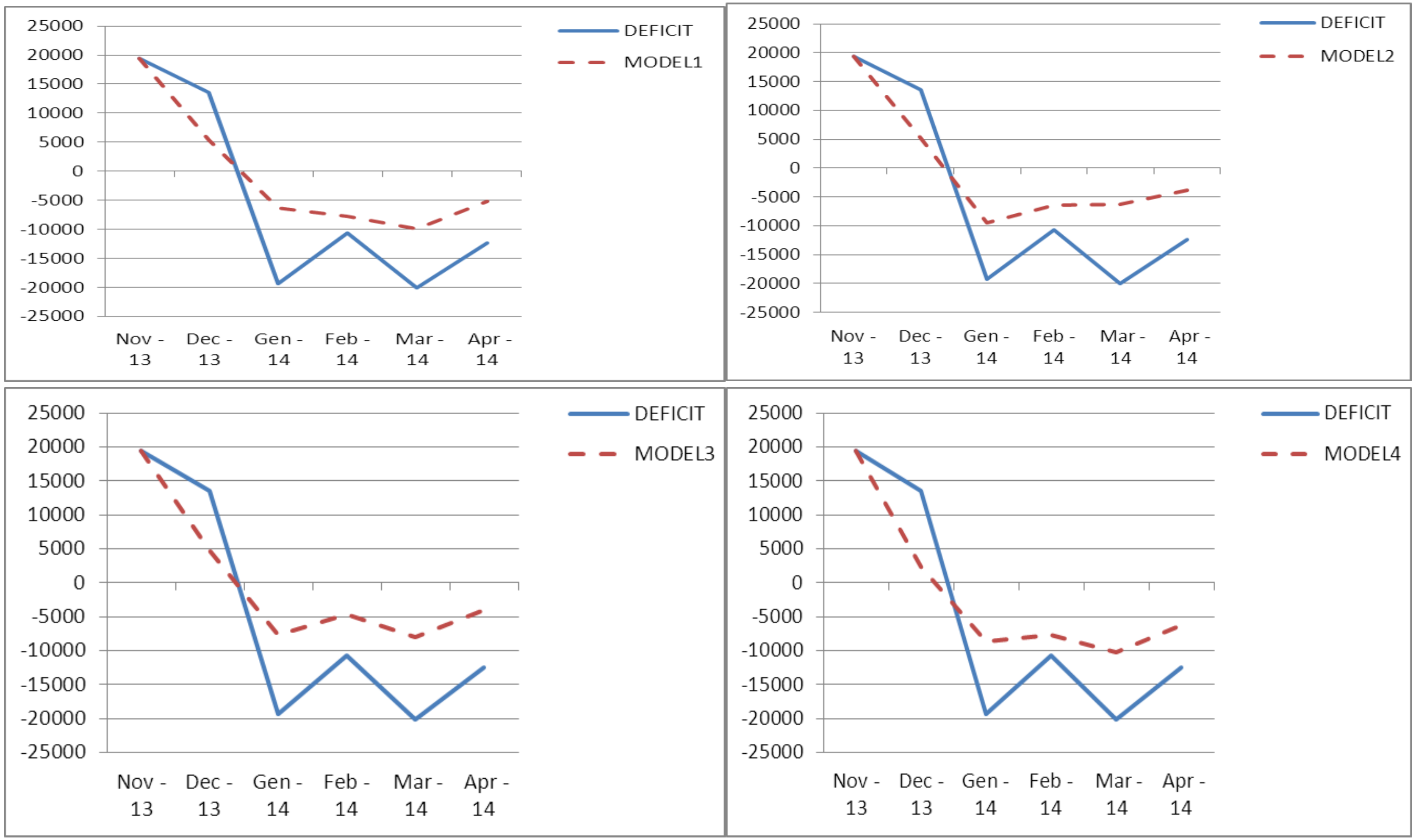
Table IV.4. Rolling windows estimation

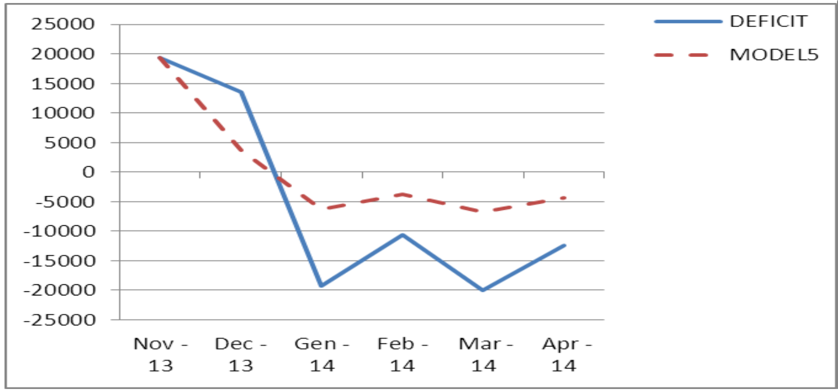
Estimation			Nowcast	
31/12/1992	31/05/2013		31/06/2013	31/10/2013
31/12/1992	31/06/2013		31/07/2013	31/11/2013
31/12/1992	31/07/2013		31/08/2013	31/12/2013
31/12/1992	31/08/2013		31/09/2013	31/01/2014
31/12/1992	31/09/2013		31/10/2013	31/02/2014
31/12/1992	31/10/2013		31/11/2013	31/03/2014
31/12/1992	31/11/201		31/12/2013	31/04/2014
31/12/1992	31/12/2013		31/01/2014	31/05/2014

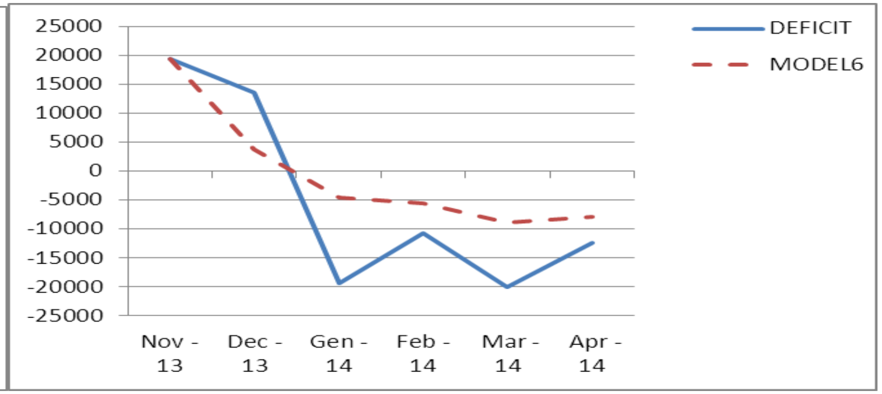
Note: The models are estimates from 31/12/1992 to 31/05/2013 and then predicted As for estimation and predicting time, the time range is from 31/06/2013 to 31/10/2014 and moves over rolling windows of one month.

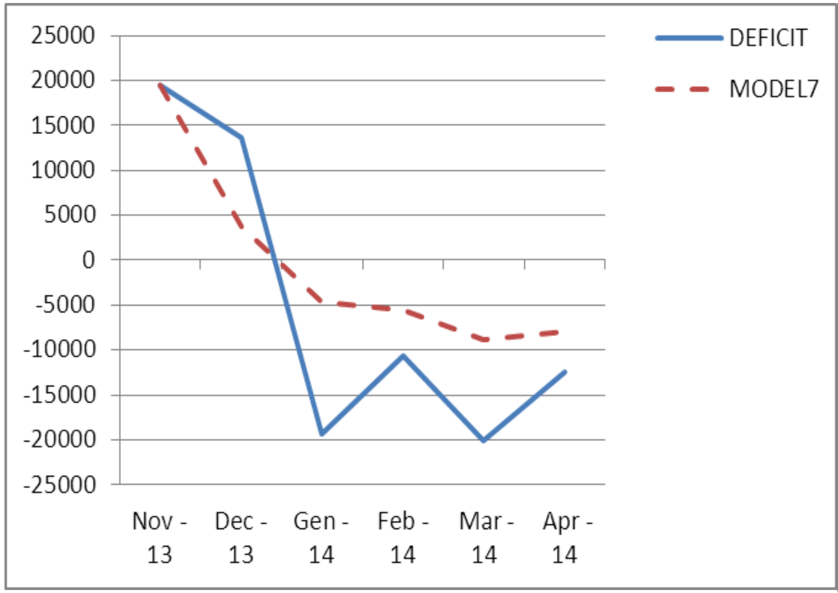
All the forecast models of the deficit are in the full sample, numbered from 1 to 14, shown in Figure IV.4. As the graphics show, they are accurate in predicting the direction of change in the deficit variable. The continuous lines are the observed deficit and the discontinuous lines are the models in the forecast time from 31/12/2013 to 30/4/2014. As the figures show, the models start at the same point and follow the trend of the observed deficit.. When change occurrs in the time series the models are able to predict it. For example, the observed deficit decreases from November 2013 and starts increasing at the beginning of 2014. In February, it begins decreasing again and in March 2014 I starts to increase. The models are able to capture the trend of the deficit and each single change in its tendency. In particular, models 1 to 10 are able to capture this behavior and models 11,12 and 13 are closer in terms of prediction to the deficit values. Indeed, these latter models seem to be the most accurate.

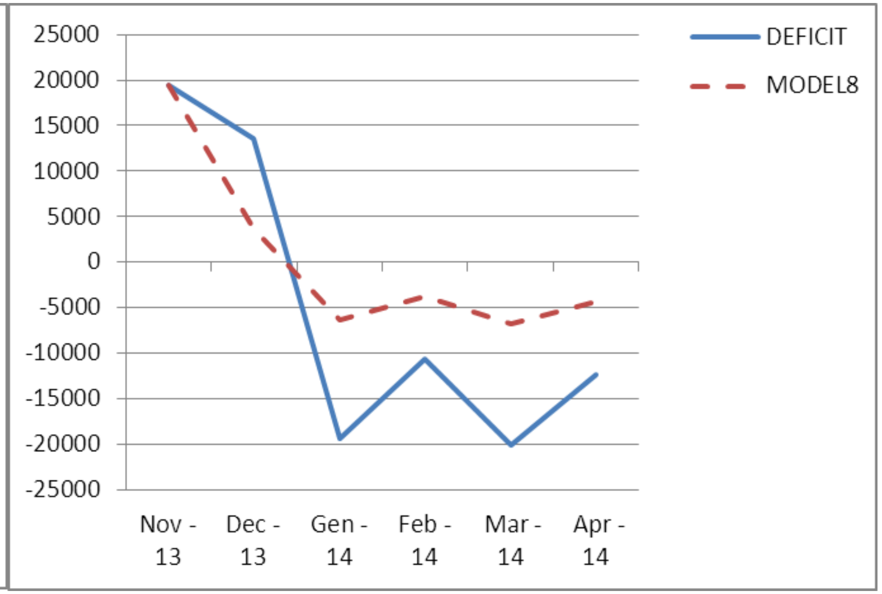
Figures IV.4. Comparison of deficit evolution and forecasts by competing models

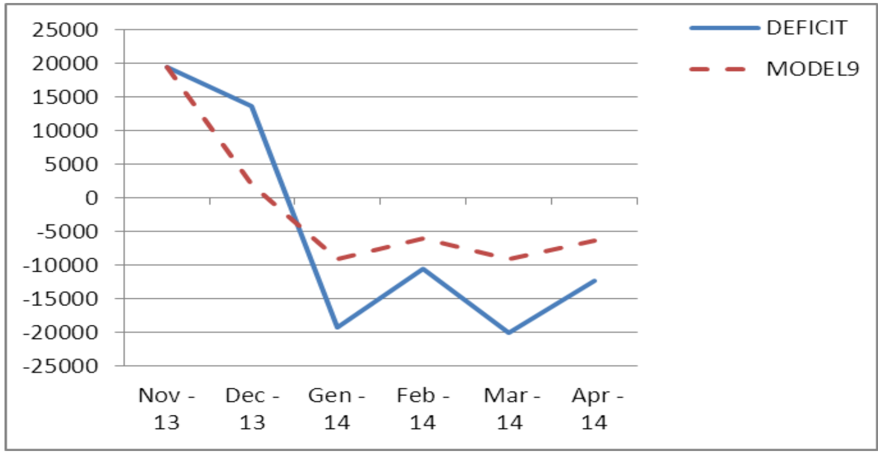


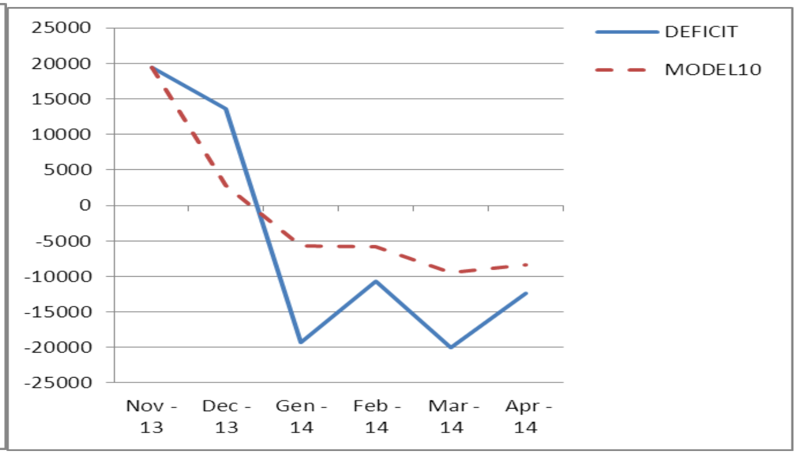

25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL5

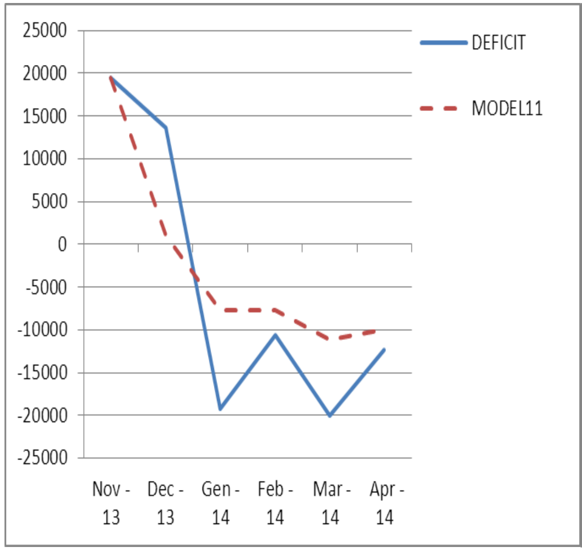

25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL6

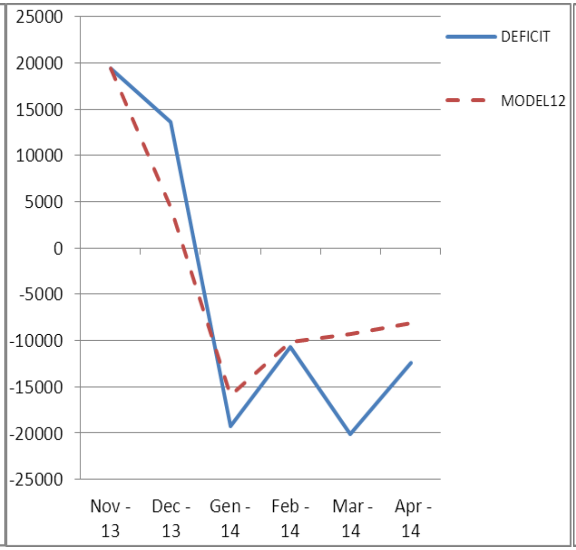

25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL7

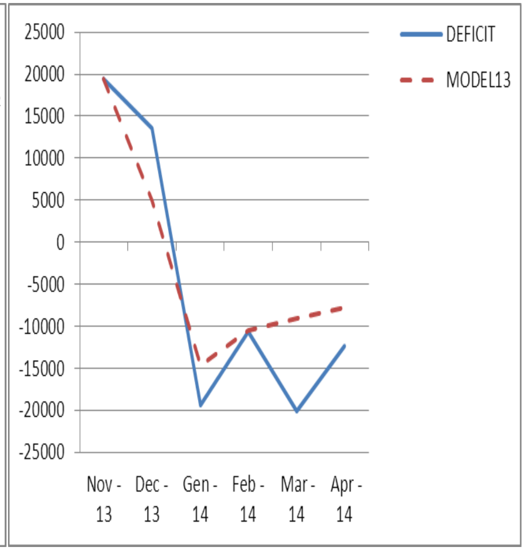

25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL8


25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL9


25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL10


25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL11


25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL12


25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
-25000
Nov - 13
Dec - 13
Gen - 14
Feb - 14
Mar - 14
Apr - 14
DEFICIT
MODEL13

IV.*3.3. Post-evaluation of nowcast: no parametric test*

In this Chapter, the accuracy analysis is provided via a qualitative test instead of a quantitative analysis[15], in the form of a non-parametric test. In this case, the deficit's monthly volatility is analysed in terms of its trend, and the qualitative test provide the best way to accurately understand if the models have the same pattern. Indeed, this forecasting measure focuses on whether the indicators are accurate in predicting the direction of the change in the target variable under consideration (Perez, 2007). In this case, I provide the percentage of correctly signed predictions and the Wilcoxon signed rank test of directional accuracy. This test compares the median of a single column of numbers against a hypothetical median. This test is used to evaluate the difference between the magnitudes and signs of paired samples to assess whether their population mean ranks differ. Originally proposed in 1945 by statistician Frank Wilcoxon (1945), the Wilcoxon signed rank test is used to produce a null hypothesis in cases where the population does not conform to normal distribution.

IV.4. Results of the nowcasting performance

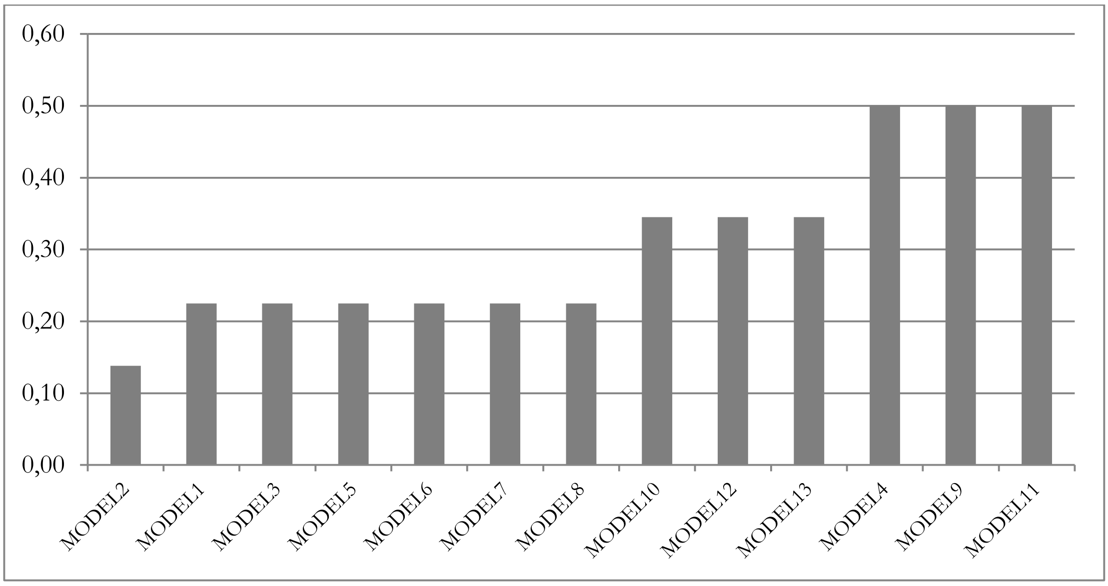
I compute the Wilcoxon test for the full sample and over recursive time to assess the performance of the models over time.

IV.*4.1. Wilcoxon test full sample*

The monthly models with the full sample consider the estimation time 1992:01-2013:12 and the predicting time 2013:12-2014:5. The results of the Wilcoxon signed rank test show good nowcasting performarmance for all of my models. Indeed, the p-values in the figure are positive and between 0.1 and 0.5. These results show that each model is able to predict the dynamic of the deficit in terms of accuracy between the models.

[15] The high levels of the MAPE and the Diebold and Mariano test (1995) from 30% to 50% in all the models show the inefficiency, in this case, of this quantitative analysis.

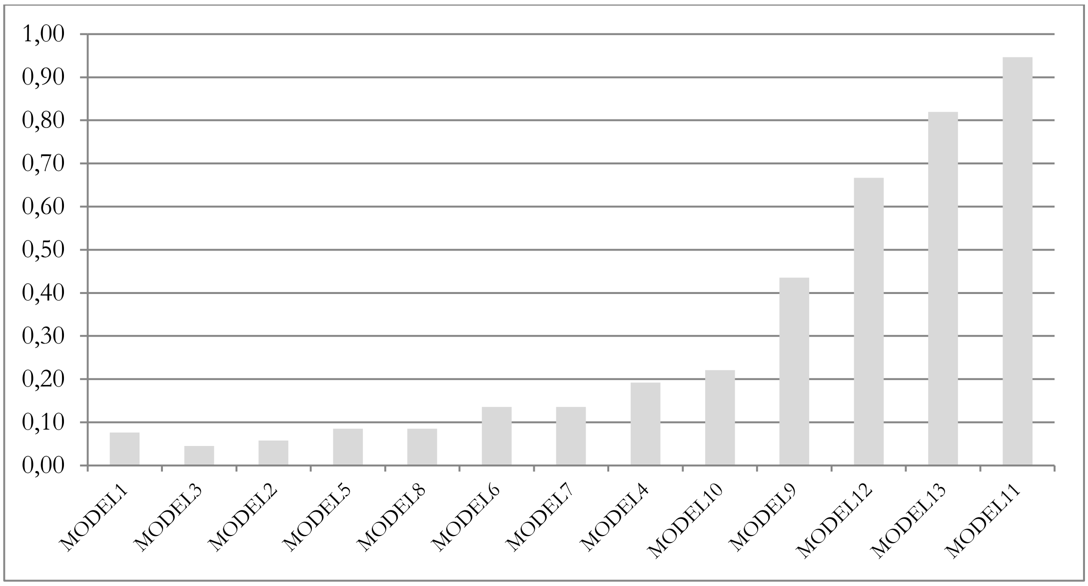
Figure IV.5. Wilcoxon test full sample models



IV.*4.2 Wilcoxon test recursive models*

We analyse the recursive models that estimate the monthly deficit during the last year of the sample with rolling windows of one month. This exercise allows us to examine the accuracy performance over time. For this purpose I compute the models by estimating over the recursive scheme from 1992:01 to 2013:5 and making predictions from 2013:06 to 2013:10; then I move one month in the estimate, as in the prediction, and I compute the models from 1992:01 to 2013:6, and predict from 2013:07 to 2013:11 and so on until the study has been filled in for the whole year. In this case, the Wilcoxon test produces p-values that are all positive, and this indicates that all of the recursive models are able to predict the deficit's trend. The test also shows high differences among the models. In particular, Figure IV.6 shows that the Linear Regression Models (models 11,12 and 13), which they consider in their equations the economic and fiscal variables and also the Economic Confidence Index and Google trends. The Linear Regression Models seem to improve the accuracy compared with the other models. Indeed, the values of the recursive models are between 0,6 (model 12), 0.85 (model 13) and 0.96 (model 11). As conclusions, the accuracy improves for these three models wich registered Wilcoxon test with the highest values. I can reject the null hypothesis that the difference between the median of the models is zero, and and can conclude that they have the best nowcasting accuracy.

Figure IV.6. Wilcoxon test recursive models



IV.5. Conclusions

Deficit forecasts are published by national and international agencies once or twice a year. The purpose of this work is to nowcast the deficit by using monthly data from the last two decades in Italy to provide real-time information on those months for which official forecasts are not provided. To this end, in this Chapter, I employ different models, such as time series (AR, ARMA, ARMAX), Vector autoregressive (VAR), and linear regression models (LM) using monthly data on public finance, the economics' confidence index, and Google Trends. I obtain the government deficit trend using the full sample for estimation and nowcasting the last month, and I perform a recursive analysis of the models to assess the model's stability over the last year of the sample. Evidence from the non-parametric test (the Wilcoxon test) shows that parsimonious models improve in forecast accuracy and linear regression models outperform both time series models as the VAR.

The models computed could provide a neutral and transparent assessment of the of the consistency between the observed budgetary data and the official data. The models also provide a tool to generate ex-ante corrective actions on budgetary plans. European-level monitoring of the monthly deficit could improve the control of public finance and the quality of the recommendations made by European Commission to Italy.

IV.6. References

Askitas, Nikos & Zimmermann, Klaus F., 2015. *The Internet as a Data Source for Advancement in Social Sciences*, IZA Discussion Papers 8899, Institute for the Study of Labor (IZA).

Banbura, M., Giannone, D., & Reichlin, L. (2010). *Nowcasting* (CEPR Discussion Papers No. 7883).

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2012) Now-Casting and the real-time data flow (Universitee Libre de Bruxelles, Working paper No. 26).

Banerjee, A., Marcellino, M., & Masten, I. (2006). Leading indicators for euro area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics*, this issue.

Beetsma R., Giuliodori M., Walschot M., & Wierts P., From budgetary forecasts to ex post fiscal data: Exploring the evolution of fiscal forecast errors in the European Union. *Contemporary Economic Policy*, forthcoming.

Bianchi, G., Cesaroni, T., & Ricchi, O. (2013). Previsioni delle spese del bilancio dello stato attraverso i flussi di contabilità finanziaria. *Rivista di Politica Economica, SIPI Spa, 1*, 271-326.

Blanchard, O., & Perotti, R. (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics, 117*, 1329–1368.

Boivin, J., & Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking, 1*(3), 117-151.

Castle, L., Hendry, D.F., & Kitov, O.I. (2013). Forecasting and nowcasting macroeconomic variables: A methodological overview (University of Oxford, Department of Economics, Economics Series Working Papers No. 674).

Cimadomo, J. (2012). Fiscal Policy in Real Time. *The Scandinavian Journal of economics,* 114: 440–465.

Clements, M. P., & Hendry, D. F. (2003). *Forecasting in the National Accounts at the Office for National Statistics* (Report no 12, Statistics Commission).

D'Agostino, A., McQuinn, K., & O'Brien, D. (2011). Nowcasting Irish GDP (MPRA paper No. 32941). University of Munich, Germany.

D'Amuri, F. and Marcucci, J. (2009). *Google it! Forecasting the US unemployment rate with a Google job search index* (Bank of Italy discussion paper No. 18248).

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253–263.

Favero, C., & Marcellino, M., (2005). *Modelling and forecasting fiscal variables for the Euro Area* (Bocconi University Working Papers No. 298).

Ferrara, L., Guegan, D., & Rakotomarolahy, P. (2010). GDP nowcasting with ragged-edge data: A semiparametric modelling. *Journal of Forecasting, 29*, 186-199.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). *The MIDAS touch: Mixed Data Sampling regression models.* Mimeo, Chapel Hill, N.C.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews, 26*, 53–90.

Ghysels, E., Ozkan, N. (2012). Real-time predictions of the U.S. Federal Government Budget: Expenditures, revenues and deficits (UNC Kenan-Flagler Research Paper).

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting GDP and inflation: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55,* 665–676.

Giannone, D., Reichlin, L., & Simonelli, S. (2009). Nowcasting Euro Area economic activity in real-time: The role of confidence indicators. *National Institute Economic Review, 210*, 90–97.

Giovannini, A., (1991). L'utilizzo dell'analisi delle serie storiche per la previsione del fabbisogno di cassa del settore statale. (eds) Il Mulino, Informazioni e previsioni per la finanza pubblica.

Hendry, D. F., & Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics, 29,* 216–227.

Keereman, F. (1999). The track record of the Commission forecasts (Economic Papers No. 137).

Leal T., Pedragal, D. J., & Perez, J. J. (2009). Short-term monitoring of the Spanish Government Balance with mixed-frequencies models (Banco de España Working Paper No. 0931).

Merola, R., Pérez, J. J., (2013). Fiscal forecast errors: Governments versus independent agencies? *European Journal of Political Economy, 32(*C), 285-299.

Mitchell, J., Smith R. J., Weale M. R., Wright S., & Salazar, E. L. (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth. *The Economic Journal, 115(*501)*,* 108–129.

Moulin, L.M., Salto, A., & Silvestrini Veredas, D., (2004). Using intra annual information to forecast the annual state deficits. The case of France (Core discussion paper, 48).

Onorante, D. J., Pedregal, Pérez, J. J., & Signorini, S. (2010). The usefulness of infra-annual government cash budgetary data for fiscal forecasting in the euro area. *Journal of Policy Modeling, 32*(1), 98-119.

Pedragal, D. J., Pérez, J. J. (2010). Should quarterly government finance statistics be used for fiscal surveillance in Europe? *International Journal of Forecasting, 26*(4), 794-807

Pérez, J. J. (2007) Leading indicators for euro area government deficits. *International Journal of Forecasting 23*(2), 259-275.

Pina, A.M., & Venes, N.M. (2011). The political economy of EDP fiscal forecasts: An empirical assessment. *European Journal of Political Economy, 27*(3), pp. 534-546.

Perotti, R. (2002). *Estimating the effects of fiscal policy in OECD countries.* Mimeo, European University Institute.

Pozzuoli, S., & Ercoli, R., (2008). *Quanto sono credibili le previsioni di finanza pubblica di medio e lungo termine in Italia? Un modelo previsionale del fabbisogno del settore statale* (SIEP Working paper).

Smets, F., Warne A., & Wouters, R. (2013). *Professional forecasters and the real-time forecasting performance of an estimated new keynesian model for the Euro Area* (European Central Bank, Working Paper No. 1571).

Stock, J. H., & Watson, M. W. (2002). *Combination Forecasts of Output Growth and the 2001 Recession*. Mimeo, Harvard University and Princeton University.

Wilcoxon, F., (1945). Individual comparisons by ranking methods. *Biometrics Bulletin 1*(6): 80–83.

ANNEX IV.1. The Wilcoxon test

In the Wilcoxon test, the assumptions are as follows:

1. Data are paired and come from the same population.
2. Each pair is chosen randomly and independently.
3. The data are measured at least on an ordinal scale, but need not be normal.

Let N be the sample size (the number of pairs). Thus, there are a total of $2N$ data points. For $i = 1, \ldots, N$ let $x_{1,i}$ and $x_{2,i}$ denote the measurements. With H_0: the median difference between the pairs is zero and with H_1: the median difference is not zero.

1. For $i = 1, \ldots, N$, calculate $|x_{2,i} - x_{1,i}|$ and $sgn\ (x_{2,i} - x_{1,i})$, where sgn is the sign function.
2. Exclude pairs with $|x_{2,i} - x_{1,i}| = 0$. Let N_r be the reduced sample size.
3. Order the remaining N_r pairs from smallest absolute difference to largest absolute difference, $|x_{2,i} - x_{1,i}|$.
4. Rank the pairs, starting with the smallest as 1. Ties receive a rank equal to the average of the ranks they span. Let R_i denote the rank.
5. Calculate the test statistic W, the absolute value of the sum of the signed ranks: $W = \left|\sum_{i=1}^{N_r}[sgn(x_{2,i} - x_{1,i}) * \ R_i]\right|$.

   As N_r increases, the sampling distribution of W converges to a normal distribution. Thus, for $N_r \geq 10$, a z-score can be calculated as

$$z = \frac{W-0{,}5}{\sigma_W}\,, \sigma_W = \sqrt{\frac{N_r(N_r+1)(2N_r+1)}{\sigma}}$$

If $z = z_{critical}$ then reject H_0 For $N_r < 10$, W is compared to a critical value. If $W \geq W_{critical}$, N_r then reject H_0

Annex IV.2. Results of Wilcoxon test

Table A.IV.2.1. Wilcoxon Test: Results nowcasting full sample

Model 1				Model 2			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	2	7.5
negative	4	12	7.5	negative	4	13	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.214			z	-1.483		
Prob > z	0.2249			Prob > z	0.1380		
variance	13.75			variance	13.75		

Model 3				Model 4			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	5	7.5
negative	4	12	7.5	negative	4	10	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.214			z	-0.674		
Prob > z	0.2249			Prob > z	0.5002		
variance	13.75			variance	13.75		

Model 5				Model 6			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	3	7.5
negative	4	12	7.5	negative	4	12	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.214			z	-1.214		
Prob > z	0.2249			Prob > z	0.2249		
variance	13.75			variance	13.75		

Model 7				Model 8			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	3	7.5
negative	4	12	7.5	negative	4	12	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.214			z	-1.214		
Prob > z	0.2249			Prob > z	0.2249		

Model 9				Model 10				Model 11			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	4	7.5	positive	1	5	7.5
negative	4	12	7.5	negative	4	11	7.5	negative	4	10	7.5
zero	0	0	0	zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15	all	5	15	15
z	-0.674			z	-0.944			z	-0.674		
Prob > z	0.5002			Prob > z	0.3452			Prob > z	0.5002		
variance	13.75			variance	13.75			variance	13.75		

Model 12				Model 13			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	4	7.5	positive	1	4	7.5
negative	4	11	7.5	negative	4	11	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-0.944			z	-0.944		
Prob > z	0.3452			Prob > z	0.3452		
variance	13.75			variance	13.75		

Table A.IV.2.2. Wilcoxon Test: Results nowcasting recursive sample

Model 1				Model 2			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	2	7.5
negative	4	12	7.5	negative	4	13	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.21			z	-1.48		
Prob > z	0.22			Prob > z	0.13		
variance	13.75			variance	13.75		
Model 3				Model 4			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	5	7.5
negative	4	12	7.5	negative	4	10	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.21			z	-0.67		
Prob > z	0.22			Prob > z	0.50		
variance	13.75			variance	13.75		

Model 5				Model 6			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	3	7.5
negative	4	12	7.5	negative	4	12	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.21			z	-1.21		
Prob > z	0.22			Prob > z	0.22		
variance	13.75			variance	13.75		
Model 7				Model 8			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	3	7.5
negative	4	12	7.5	negative	4	12	7.5
zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15
z	-1.21			z	-1.21		
Prob > z	0.22			Prob > z	0.22		
variance	13.75			variance	13.75		

Model 9				Model 10				Model 11			
sign	obs	sum ranks	expected	sign	obs	sum ranks	expected	sign	obs	sum ranks	expected
positive	1	3	7.5	positive	1	4	7.5	positive	1	5	7.5
negative	4	12	7.5	negative	4	11	7.5	negative	4	10	7.5
zero	0	0	0	zero	0	0	0	zero	0	0	0
all	5	15	15	all	5	15	15	all	5	15	15
z	-0.67			z	-0.94			z	-0.67		
Prob > z	0.50			Prob > z	0.34			Prob > z	0.50		
variance	13.75			variance	13.75			variance	13.75		

Model 12			
sign	obs	sum ranks	expected
positive	1	4	7.5
negative	4	11	7.5
zero	0	0	0
all	5	15	15
z	-0.94		
Prob > z	0.34		
variance	13.75		

Annex IV.3. Autocorrelations and partial autocorrelations of deficit

Figure A.VI.3.1. AC – Autocorrelations of deficit

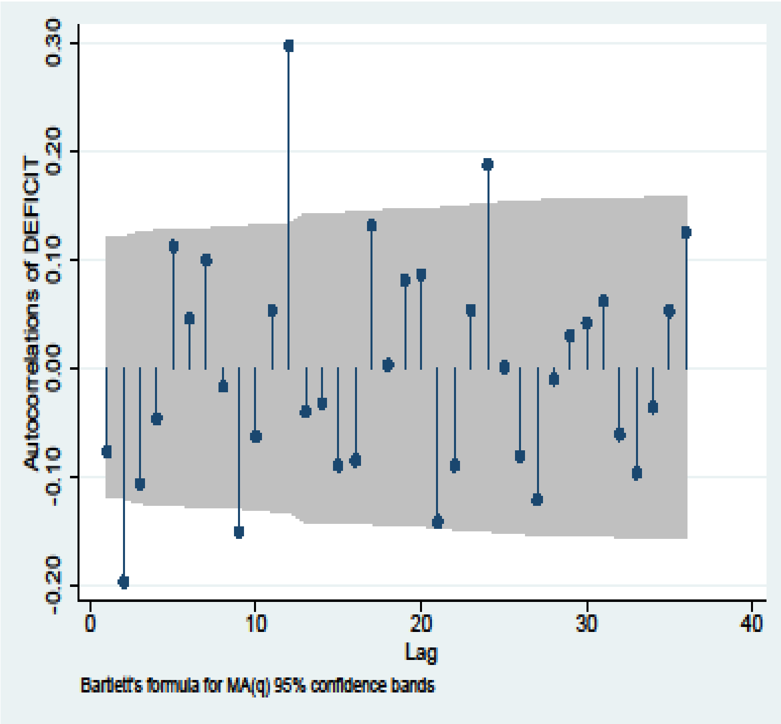
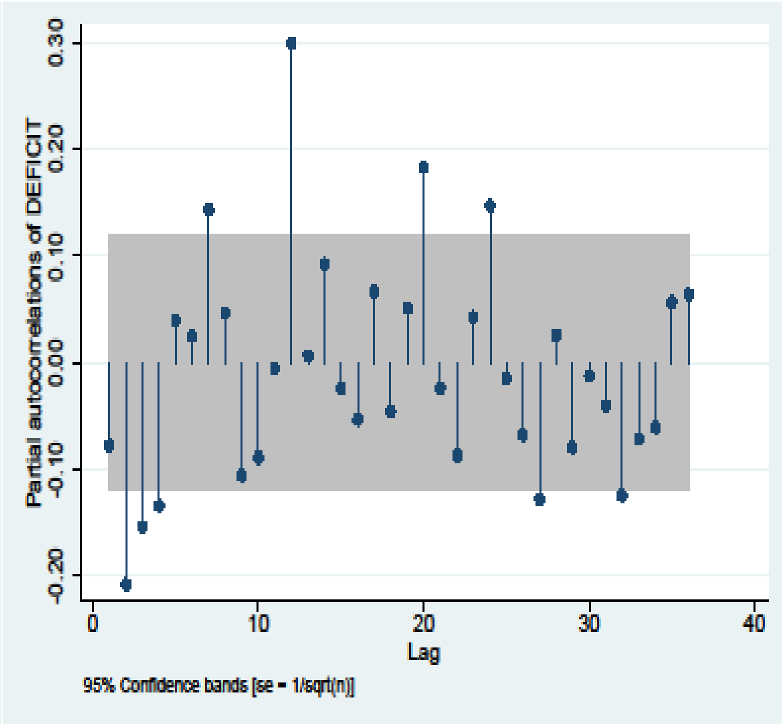


Figure A.VI.3.2. PAC – Partial autocorrelations of deficit



Chapter V.

Conclusions

V.1. Main contributions and overall view

This thesis has examined several aspects of public finance forecasts in Italy over the last two decades. In this final Chapter, I briefly discuss the main findings, the results of the analysis, the policy implications, and the lines of research that this work has inspired to me.

This dissertation has centered its analysis on two specific aspects of public finance forecasts: firstly, on evaluations of the quality of the Italian deficit/GDP forecasts made by different agencies (Chapter II); secondly, on experimenting with different approaches to the aim of improving the quality of the forecasts through: methods of combination (Chapter III) and using high frequencies variables for nowcasting (Chapter IV).

Considering the institutional and economic context from March 2012 through to the introduction of the European Semester, the EC provides recommendations to individual member states that are based less on outcomes and more on forecasts. With a preventive vision, it seems to be necessary to collect data of the most important economic and financial variables at European level. All member states have committed to achieving the Europe 2020 targets and have translated them into national targets.

From this political and economic scenario and the literature about the state of the art on forecasts, I tried to reflect on the first question that inspired Chapter II: "Are fiscal forecasts for Italy reliable enough in order to plan economic and political strategies for the future?" To answer this question, the first step was to analyse the deficit/GDP errors in the published national, public and private, and international forecasts during the last two decades, and then to identify which of them adapted their predictions better to the financial and economic crisis. The second step was to investigate during which months of the year better forecasts were made. This is important because each year the Commission undertakes a detailed analysis of EU Member States' plans for budgetary, macroeconomic, and structural reforms and provides them with recommendations for the next 12-18 months. In this context, the tests of data accuracy released by five private, three international, and one national forecaster from the years 1992 to 2012 show that private sector deficit/GDP forecasts paint the best picture about a country's public deficit situation, and that the forecast data provided in December of the previous year is the most accurate. Econometrically, the forecasts are efficient, unbiased, and free of serial correlation when taken individually, as a whole, and as a panel.

In this sense, Chapter II's results show that forecasts' accuracy increases when they are produced by independent agencies -- a result also found by Frankel and Schreger (2013) -- and also when the data became available (when the projection horizon became shorter). This is the case for the forecasts provided during the current year, and in December in particular. Even though the quantitative analysis shows worse performances for the year ahead horizons, as for the current horizons, the qualitative analysis indicates efficiency and unbiasedness. In general, all forecasters tend to under-predict the deficit when the economy is slowing down, especially during recessions, and after that, they over-predict. The explanation of this point is provided by some evidence in the literature (Merola and Perez, 2012), in which it is demostrated that GDP errors influence the government deficit, in the way that a negative growth shock produces ex-post optimistic revenue and deficit forecasts.

The results of this chapter suggest that the accuracy of the deficit/GDP forecasts published by private national agencies is statistically different from that of the public national and international agencies for the horizons of current year and year ahead. All nine agencies improve on the accuracy provided by a naïve model for both horizons. The results are broadly in line with the literature, particularly Artis and Marcellino (1998, 2001), Keerman (1999), Koutsogeorgopoulou (2000), and Abreu (2011). These authors demonstrate that all of the forecasts provided by national and international agencies are accurate, unbiased, and efficient.

The overall picture of these findings demonstrates that the magnitude of the forecast errors, the institutional nature of the agencies that provide public finance forecasts, the timing of publication and the time of prediction, and the deviations that the forecasts assume during the economic crisis are absolutely necessary in order to assess forecasts' usefulness. This Chapter's research provides some interesting results for policy makers: firstly, it could be useful to monitor the "current year" fiscal forecast variable during the year, considering the high level of accuracy provided by private agencies. Secondly, in the assessment of the EC during the European Semester, the forecast data provided in December during the previous year "year ahead" is a significant indicator of the future trend of that variable.

The following two chapters both respond to the question "Is it possible to improve the accuracy of fiscal forecasts?" With this question in mind, the study aims to improve the accuracy of public deficit forecasts by using combination techniques (Chapter III) and by nowcasting (Chapter IV). In particular, in Chapter III, I try to develop a different way of combining a large number of forecasts from both private and public forecasters. The way I do this is by averaging forecasts from different sources in a variety of ways. I include simple as well as more advanced averaging techniques that account for past forecasting performance to compute a combined forecast. I prove that the combined forecasts from national, international, and private agencies of the deficit/ GDP ratio for Italy over the period from 1992-2012 do better than any single forecast. There are many factors supporting the thesis that combining forecasts works well. The main findings of empirical works show that the accuracy of combining forecasts is higher than that of the individual forecasts (Makridakis, 1989).

The reasons are the following:
1) Measuring the wrong thing: there are many macroeconomic variables that are endogenous to the economic cycle. This means that to forecast these variables, the methods that are used for forecasts use other variables to explain the former. These proxies introduce systematic biases to the measurement of the real value, which reduces the forecasts' accuracy. In this case, combining reduces the risk of bias.
2) Measurement errors: in any forecast there is an error, no matter which model is used. In general, it is important to compute the size of the error and determine if it is systemic or not. Also, there are changes in accounting and in the manner in which forecasters download the data, and the definitions of the variables are not always the same. Combining works because it averages such errors.
3) Unstable changing patterns or links: the models tend to consider the link between variables and the criteria as though they are constant over time. It is evident that this is not possible in the real world, where events, crises, and

situations change all the time, and change continuously and synchronically the link with the variables. Combination levels out this instability link between variables.
4) Models minimize past errors: there is evidence about the strict serial correlation with the errors and the presence of bias that depends on how well models minimize year ahead forecasting errors when tested against available data. Combination tends to avoid the selection of the best model by this process.

A huge literature finds that the practice of combination improves the performance of single models, for example, Timmermann (2006), with a large set of fiscal variables, Artis and Marcellino (2002, 2004) on the deficit ratio for G7 countries, and Ozkan (2011) on the US deficit ratio from 1970 to 2005. Following this relevant practice, I have performed many accuracy tests to assess the quality of the simple and more complex combination models. As in Chapter II, I have used, for example, the Theil test (1971) (compare the RMSE of a single and a combination model with a RMSE of a naïve model) and the Diebold-Mariano test (1995) (compare the MSE of each single and combination model to each other and with a naïve model). In this computation, I use the entire full sample (1992-2012). In practice, however, predictive accuracy changes over time. Finding an indicator that predicts well in one period is no guarantee that it will predict well in later periods. To investigate the presence of instability over time in the performances of each single and combined model, I used a specific test, the fluctuation test (Giacomini and Rossi, 2010). In particular, this test analyzes the evolution of the relative performance in the sample. In particular, it tests fluctuations in the relative predictive performance of forecasting methods over time by comparing the MSE provided by two different models, over rolling windows (10 years).

The main results of the accuracy analysis show that combination models perform better than single forecasts both for current year and year ahead. The best models seem to be weighted forecast combination and Rbest. The first one is computed in an advanced way by running the regression linera models. The second one is the discounted mean square forecast error; it gives each individual predictor a weight that is inversely proportional to the predictor's mean square forecast error. A discount factor is applied to attach greater weight to the recent predictive ability of the individual predictor. In this case, Rbest considers only the last best 4 periods of the sample. The overall view of these findings demonstrates that applying nine different combination techniques to the current year and year ahead national (private and public) and international forecasts of the Italian budget deficit over the period from 1992 to 2012 results in substantial gains in forecast accuracy. In particular, the predictive power of the weighted forecast combination models (WFC and Rbest) are better than others and, in terms of stability, they are structurally stable enough to be better than a naïve model over time. The policy implications of these finding demonstrate that weighted forecast combination models present the right way to improve the fiscal forecasts published during the year. The goal of further analysis could be to find new "weights" for the combination models to improve fiscal forecast accuracy and its stability over time.

In Chapter IV, I have experimented with other techniques to improve the performance of forecasts. The method used is nowcasting with the use of high frequency variables. The basic idea is to nowcast the deficit by using monthly data from the last two decades in Italy to provide real-time information for those months in which official forecasts are not published. I do this by running econometric models using specific variables as fiscal variables, Economic Confidence Index and Google trends (D'Amuri, 2009, Vosen and Schmidt, 2011). I include simple and parsimonious models to compute new forecasts. First, I consider different models of autoregressive (AR) and autoregressive moving average (ARIMA), which many studies have considered to have a good performance in macroeconomic applications (Marcellino, 2004a, 2005a; Stock and Watson, 2002, Banerijee, Marcellino and Masten, 2006, Favero and Marcellino 2005).

Second, I use a Vector Autoregression Model (VAR) to model macroeconomic variables and forecast, as in Favero and Marcellino (2005). Third, I estimate regression models through the Linear Regression Model (LRM) using the variables from Favero and Marcellino (2005) and Ghysels and Ozkan (2012). I nowcast the deficit, following these two prediction's schemes: full sample, computing the estimation in all the sample and nowcast the last four months and recursive sample, computing estimation until the last year of the sample and recursive nowcast until the last year. Even if nowcasting (as it is defined in the literature and as I treat it in this dissertation) is a recent field for academic research, many studies support the validity of this method. For example, Giannone et al. (2008, 2009) use the factor model for GDP USA and Europe; and Bánbura et al. (2010, 2012) use the dynamic daily factor model for GDP.

Applying the method of nowcasting to the deficit is innovative in this context, according two approaches. Firstly, it is methodologically innovative in the sense that applied techniques for nowcasting the public deficit use parsimonious models that are replicable and simple, so that they can be used without the use of endogenous variables. Secondly, providing an original database. These databases collect high frequency data about public finance, Economic confidence index, and Google Trends available at the same time with the aim of predicting the deficit. Indeed, these two approaches add to the recent literature by using econometric models that are applied for nowcasting public finance, and also in the databases used for this goal. Only a few studies have tested the topic, such as Ghyselle and Ozkan (2012), using ADL/MIDAS for expenditure, revenue and deficits in the USA.

In conclusion, nowcasting the Italian budget deficit over the period from 1992:12 to 2014:5 by applying different models shows that parsimonious models improve forecast accuracy. In particular, the predictive power of the simple regressions (with economic, confidence index, and Google Trend variables) is better than the others. The main finding is that all the models can capture the future trend of the actual deficit. In particular, the Linear Regression Models are the most accurate at forecasting increases or decreases in the monthly deficit. My analysis is useful for policy makers during the fiscal year; specifically it provides the forecast deficit trend in those months in which official forecasts are not published. The policy implication of this work could be that with this real-time budget indicator, budget plans could be adjusted and updated on a monthly basis. It is therefore most useful for revisions due to new information, in the run-up to the European Semester, and to inform Parliament on budget developments. Projections based on the real-time forecast can help to focus the political discussion on the government's budget plans. Its indications can aid in avoiding a spiralling of debt and a fiscal crisis.

Many insights could be offered to policy makers in this latest work. Indeed, the models computed could provide a neutral and transparent assessment of the consistency of the observed budgetary data with the official data published one or twice a year. The described approach provides a tool to generate ex-ante corrective actions on budgetary plans; the European-level consequence is a monitoring of the monthly deficit to improve the public finance control and the policy recommendations from European Commission to Italy. This conclusion inspires different ideas for further researche. One way forward could be to link the high frequency now-casting framework with a quarterly structural model through the bridge equations and MIDAS (Perez, 2007, Giannone et al., 2014). In this way, it could be possible to predict deficits from real-time data. The aim of doing this exercise is to assess if the forecasts provided by parsimonious models outperform the forecasts from the OECD, the IMF, the EC, the national government, and private agencies. Another idea could be to follow the work of Ghyselle and Ozkan (2012) and nowcast the deficit while considering revenues and expenditures separately. The reason for this experiment could be to identify, with greater accuracy, if the forecasts' errors are due more to the revenues or expenditures, and then to correct the models. Another idea could be to add to the estimate in the full sample and the recursive sample, also applying it at a rolling windows scheme. In this way, it could be possible to have an updated prediction of the deficit each month, which would be useful for monitoring the monthly deficit at the country and European level. An overall view of the three chapters taken as a whole reveals that the quality of the forecasts relating to the variable annual deficit / GDP is able to ensure efficient performance.

The combination of these forecasts considerably improves their performance, even compared to a naïve model. Even if financial crises and sudden changes in the economic cycle are difficult to predict, the work of this thesis shows that by analysing forecasts' quality, independent agencies make a valuable contribution in support of the national and international government publications. Also, the time of year in which the forecasts are published is important. In the final months of the year, forecasts are less error-prone for both the current year and the year ahead, even during the crisis of 2001 and 2008. In addition, this study shows that when the economy registers a shock, the weighted combination of the forecasts and the contextual introduction of it into the naive model, is able to quickly adapt to the performance of each single model and provide better forecast performances in general. This thesis also shows that the introduction of monthly variables is able to capture these immediate changes in the economic and financial system and improve the ability of deficit forecasts to obtain satisfactory results regarding the trend, although the series is significantly volatile.

V.2. Discussion on further developments

In general, this thesis has provided deep reflections and opened space for a discussion on further developments.

The first issue to consider is the applicability of this work to Italy. Nowadays, legislation requires that the models on which the forecasts are based have to be published. Such transparency in the methodologies used by both the public and private forecasters could be useful to exploit and analyze in detail the various models, the variables used, the different problems of endogeneity, and also the correlation between their errors. It is known that there is an influence among forecasters that produces correlation in the errors. Studying their models in detail will make it possible to reduce these errors and improve the quality of the forecasts.

Another aspect would be to reduce the aggregation of models, replicating the analysis for revenue and expenditures in the public budget. This aggregation should take into account the composition of the budget into categories, titles, and chapters.

The second area for future investigation is related to the extension of this work to other countries of the European Union. In particular, the analysis of accuracy and combinations, as well as the production of monthly forecasts could be carried out for each member state in order to identify what is the state of their forecasts' performance, and in this way try to improve them. This analysis could also make possible a comparison between the different country forecasting models in order to capture positive predictive capabilities that could be adapted to the other models of other countries. To this end, I could also analyze patterns of forecasts in each country and compare them with the results of this study to determine if and where the forecasts could be improved. I could analyze if there is a contagion effect between countries and the forecasts produced. I could identify if there is a correlation between the errors of the various countries that could lead to a spiral of improvement of the forecasts at the European level.

The third area for further research relates to expandability in time. In particular, from the introduction of European Semester, member states rearranged the public balance according to the recommendations of the EC, such that it would no longer be based on outcomes but on forecasts. Since the year 2012, due to the innovative rethinking of this new procedure, which considers the optimization of public management in advance to avoid economic crises, the forecasts produced by the practitioner Italians have probably improved. The techniques used by different forecasters, thanks to this new procedure, have become probably more stringent and efficient. With this in mind, as a possible further investigation, it could be useful to expand this study in terms of the time horizon and examine ex-post the effect of the new procedure to analyze if the European Semester has influenced and improved the public balance and its forecasts.

V.3. References

Abreu, L. (2011). *International Organizations' vs. private analysts' forecasts: an evaluation* (Banco de Portugal working paper No. 20).

Artis, M., & Marcellino, M. (1998). Fiscal solvency and fiscal forecasting in Europe (CEPR Discussion Paper No. 1836)

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: the track record of the IMF, OECD and EC. *Econometrics Journal,* 4, 20-36.

Banbura, M., Giannone, D., & Reichlin, L. (2010). *Nowcasting* (CEPR Discussion Papers No. 7883).

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2012). Now-casting and the real-time data flow (Universitee Libre de Bruxelles Working paper No. 26).

Banerjee, A., Marcellino, & M., Masten, I. (2006). Leading indicators for euro area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics.*

D'Amuri, F. & Marcucci, J. (2009). *Google it! Forecasting the US unemployment rate with a Google job search index* (Bank of Italy Discussion paper No. 18248).

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.

Favero, C., & Marcellino, M., (2005). Modelling and forecasting fiscal variables for the Euro Area (Bocconi University Working Papers No. 298.)

Frankel, J. A., & Schreger, J. (2013). Bias in official fiscal forecasts: Can private forecasts help? mimeo, Harvard.

Giannone, D., Reichlin, L., & Small, D. (2008) Nowcasting GDP and inflation: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55,* 665–676.

Giannone, D., Reichlin, L., & Simonelli, S. (2009). Nowcasting Euro Area economic activity in real-time: The role of confidence indicators. *National Institute Economic Review, 210,* 90–97.

Ghysels, E., Ozkan, N. (2012). Real-time predictions of the U.S. Federal Government Budget: Expenditures, revenues and deficits (UNC Kenan-Flagler Research Paper).

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics, 25*, 595–620.

Giannone, D., Monti, F., & Reichlin, L. (2014). Exploiting the monthly data flow in structural forecasting (Bank of England Working Paper No. 509).

Keereman, F. (1999). *The track record of the Commission forecasts* (European Commission, Directorate-General for Economic and Financial Affairs, Economic Papers No. 137).

Koutsogeorgopoulou, V. (2000). *A post-mortem on Economic Outlook projections* (OECD Economics Department Working Paper No. 274.)

Marcellino, M. (2002). *Forecast pooling for short time series of macroeconomic variables* (CEPR Working paper No. 3313).

Marcellino, M. (2004). Forecast pooling for European macroeconomic variables. *Oxford Bulletin of Economics and Statistics, 66(*1), 91-112.

Marcellino, M. (2004a). Forecasting EMU macroeconomic variables. *International Journal of Forecasting, 20*, 359–372.

Marcellino, M. (2005a). Instability and non-linearity in the EMU in Milas C., Rothman P. & van Dijk D. (Eds), *Nonlinear Time Series Analysis of Business Cycles,* Amsterdam (forthcoming).

Merola, R., & Pérez, J. J. (2013). Fiscal forecast errors: Governments versus independent agencies? *European Journal of Political Economy, 32(*C), 285-299.

Ozkan, N. (2011). *A first step to develop a forecasting model for U.S. Government budget deficit: A regression-based method with quarterly economic indicators.* University of North Caroline at Chapel Hill, Department of Economics.

Pérez, J. J. (2007). Leading indicators for Euro Area government deficits. *International Journal of Forecasting*, *23* (2), 259-275

Stock, J. H., & Watson, M. W. (2002). *Combination forecasts of output growth and the 2001 Recession*. Mimeo, Harvard University and Princeton University.

Theil, H. (1971). *Applied economic forecasting*. North-Holland Publishing Company.

Timmermann, A. (2006). Forecast combinations, in Elliot, G., Granger, C. & A. Timmermann (Eds). *Handbook of Economic Forecasting* (136-196), North-Holland, 1, 136-196.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578.