Joint Forecast Combination of Macroeconomic Aggregates and Their Components

Cobb, Marcus P A

February 2017

Joint Forecast Combination of Macroeconomic Aggregates and Their Components

Marcus P. A. Cobb*

February 2017

Abstract

This paper presents a framework that extends forecast combination to include an aggregate and its components in the same process. This is done with the objective of increasing aggregate forecasting accuracy by using relevant disaggregate information and increasing disaggregate forecasting accuracy by providing a binding context for the component's forecasts. The method relies on acknowledging that underlying a composite index is a well defined structure and its outcome is a fully consistent forecasting scenario. This is particularly relevant for people that are interested in certain components or that have to provide support for a particular aggregate assessment. In an empirical application with GDP data from France, Germany and the United Kingdom we find that the outcome of the combination method shows equal aggregate accuracy to that of equivalent traditional combination methods and a disaggregate accuracy similar or better to that of the best single models.

Keywords: Bottom-up forecasting; Forecast combination; Hierarchical forecasting; Reconciling forecasts

JEL codes: C53, E27, E37

*The author is grateful to Andrea Carriero and Marco Mariotti for their valuable comments and support. This research was produced while studying at the School of Economics and Finance, Queen Mary University of London and the author acknowledges and is grateful for their financial support.

Non-technical Summary

Macroeconomic forecasts receive considerable attention due to the fact that different agents regularly use them in their decision making processes. In many situations the focus of attention is a particular aggregate and in others a whole set of them. This is the case at policy making institutions where understanding the dynamics underlying an aggregate forecast is important to formulate useful economic policies. In such a situation some level of disaggregation may be required.

The need for consistent forecasting scenarios means that institutions producing short-term forecasts usually rely on the bottom-up approach, that is building the aggregate forecast as the sum of its component's forecasts. The bottom-up approach however is criticized because it generally cannot approximate the underlying true process and therefore may end up being inferior in terms of aggregate forecasting accuracy than alternative methods. In this context, there is an ongoing debate on whether it is best to forecast an aggregate directly, indirectly as the sum of its component's forecasts or in a way that uses both.

The considerable effort that has been put into improving aggregate accuracy contrasts with the apparent lack of interest in disaggregate accuracy. There is some literature that devotes itself to exploiting the interdependence between components to increase their accuracy but hardly any that tries to take advantage of the benefits of forecasting an aggregate directly. The only exception we find is that of Hyndman et al. (2011) that propose a method that uses individual forecasts for all levels of aggregation and optimally reconciles them so that the outcome is a fully consistent set of forecasts. Their reconciliation method however focuses on the aggregation structure and does not take the actual forecasts into consideration.

In this paper we present a framework that benefits from both the direct and bottom-up approaches to increase overall forecasting accuracy. We do so by extending the well proven and robust results from the forecast combination literature to a setting that includes one level of disaggregation. We produce individual forecasts for the aggregate and all the components and consider them as initial guesses. Then we update them based on their relative reliability so that they comply with the identities that define the aggregate. The problem is set in a general constrained quadratic program but under fairly mild assumptions derive analytical solutions making the method very easy to apply.

Our empirical application uses GDP data from France, Germany and the United Kingdom. We find that the outcome of our method results in equal aggregate accuracy to that of equivalent traditional combination methods and a disaggregate accuracy similar or better than that of the best single models. Our results suggest that our framework successfully replicates the benefits of traditional forecast combination in terms of aggregate accuracy while increasing disaggregate accuracy by effectively imposing the aggregation structure on the individual forecasts.

1 Introduction

Assessing the state of the economy and providing an outlook for where it is heading involves interpreting large amounts of data in a way that is coherent. Macroeconomic aggregates are fundamental to this process given that they synthesise the information from countless indicators into relatively few figures. Consequently, many different people and institutions devote considerable resources to predicting key economic variables. The Survey of Professional Forecasters published by the European Central Bank (ECB, 2015)[1] serves just as an example of this.

There are many ways in which economic variables can be forecasted and when it comes to aggregates there is also the question on whether to use the disaggregate data. In this context, one line of research has centred on determining whether forecasting an aggregate as the sum of the forecasted components is better in terms of aggregate accuracy than forecasting the aggregate directly. The result of this debate, as it stands, is that it depends on the problem at hand. This view is supported by the contrasting results from many empirical comparisons[2] and the fact that for most feasible implementations which is better depends on the particular structure of the disaggregate process and the aggregation matrix (Lütkepohl, 1987).

In some applications, however, focusing solely on the aggregate is not sufficient. As Espasa and Mayo-Burgos (2013) point out, sometimes the dynamics of the components underlying the aggregate forecast are of as much interest as the aggregate itself. It is often the case that practitioners rely on methods in which the aggregate is produced as the sum of the forecasts of its components because they have to be able to explain what underlies an aggregate forecast (Esteves, 2013; Ravazzolo and Vahey, 2014). In such cases a direct approach is not a viable option.

There are strong arguments however in favour of using direct approaches when the concern is aggregate accuracy. Granger (1987) show that common factors that are relatively unimportant at an individual level may dominate the aggregate while Hendry and Hubrich (2011) argue that, given that the bottom-up strategy is usually implemented by forecasting the disaggregate components independently from each other, it cannot properly approximate the underlying multivariate process. In this context, a forecaster

[1]The ECB Survey of Professional Forecasters collects expectations on inflation, real GDP growth and unemployment in the euro area from experts affiliated with financial and non-financial institutions from within the area (Garcia, 2003).

[2]For example Espasa et al. (2002), Benalal et al. (2004), Hubrich (2005) and Giannone et al. (2014) for inflation in the Euro area; Marcellino et al. (2003), Hahn and Skudelny (2008), Burriel (2012) and Esteves (2013) for European GDP growth; and Zellner and Tobias (2000), Perevalov and Maier (2010) and Drechsel and Scheufele (2013) for GDP growth in specific industrialized countries.

that is concerned with overall accuracy would want to benefit from the direct methods if possible.

One strategy could be simply to use direct methods for the aggregate and reconcile the disaggregate forecasts. This would not take into consideration however that theoretical and empirical results suggest that both the direct and bottom-up approaches are valuable. It would be very appealing therefore to be able to benefit from both.

If the concern were only for the aggregate, a popular way of dealing with two competing forecasts would be simply to combine them. The idea of forecast combination was put forward quite a while ago in Bates and Granger (1969) and deals with the issue of exploiting in the best possible way the information contained in each individual forecasts. The literature on it is extensive and the surveys by Clemen (1989), Diebold and Lopez (1996), Newbold and Harvey (2002) and Timmermann (2006) not only give testimony of it but also highlight the robustness of the gains in forecasting accuracy due to its use.[3]

Notwithstanding the extensive literature on combination methods, almost all of it deals with one variable at a time. A notable exception is that of Hyndman et al. (2011). They propose a method that uses individual forecasts for all levels of aggregation and optimally reconciles them so that the outcome is a fully consistent set of forecasts. A striking feature of their implementation is that the combination weights depend only on the aggregation structure and not on the forecasts themselves. This apparently counter-intuitive result stems directly from a key assumption. This is that the forecast errors follow the same aggregation pattern as the data.

In this paper we present a framework that extends the notions developed in the combination literature to a setting that includes one level of disaggregation but where it is not necessary to make any assumptions regarding the forecast errors. To this effect, we approach the combination process from a slightly different perspective to that of Hyndman et al. (2011). Similarly to them, we produce individual forecasts for the aggregate and all the components and consider them as initial guesses. We then however update them based on their relative reliability so that they comply with the identities that define the aggregate.

On the one hand, our framework has the potential of increasing aggregate forecasting accuracy if relevant disaggregate information is not picked up by traditional (aggregate) combination methods. On the other hand it has the potential of increasing disaggregate forecasting accuracy by providing a binding context for the individual forecasts. The

[3] Timmermann (2006) summarize a number of rationales that make combining forecasts appealing, most of them being related to the diversification of risk. Some of these are that combining forecasts could increase the information that is used to produce the final forecast, be more robust to biases associated with misspecification and less affected by structural breaks.

gains from constraining disaggregate forecasts, at least in regards to the aggregate, are supported theoretically by Giacomini and Granger (2004).

The rest of the paper is organized as follows. Section 2 develops the framework that allows for the combination of series from two different levels of aggregation. Section 3 presents an empirical implementation using GDP data for France, Germany and the United Kingdom. Section 4 summarizes the conclusions.

2 Combining Forecasts from Different Aggregation Levels

People working on the compilation of aggregate statistics regularly face the need to balance information from different sources in order to produce official statistics. In many of those applications, like the production of national accounts and the social-accounting matrices, the reconciliation process involves a massive amount of data meaning that throughout the years automatic procedures have been proposed to iron out the differences.[4]

In a recent paper, Rodrigues (2014) cast the whole problem of balancing statistical economic data into a Bayesian framework. They suggest treating the data as stochastic processes, modelling their prior properties accordingly and finding the balanced posterior by means of relative entropy minimization.

The process proposed by Rodrigues (2014) equates to searching for a posterior distribution that is as close as possible to the prior that satisfies the required restrictions. Although their implementation is specific to balancing economic data, the principle behind their framework resembles the problem of any sort of forecast combination. The individual forecasts serve as best guesses, different forecasts have different reliability and cross-sectional identities must be met. They establish that a number of the conventional reconciliation methods are in fact particular cases of their general framework and show that there is a one-to-one correspondence. Based on this correspondence, they argue that it is possible to identify the conventional method's underlying assumptions and go on to suggest using least squares approaches when uncertainty estimates are available.

[4] Most of the methods can roughly be classified either as constrained optimization methods or as adjustment algorithms and are attributed to Sir Richard Stone (Stone et al., 1942; Stone, 1961, 1962). Both in the early days and now however advances in the automatic methods have made little impact in the actual generation of the data. Dalgaard and Gysting (2004) argue that there are technical issues but that primarily it is due to the fact that errors in primary statistics are spotted in the course of the manual process.

2.1 A Constrained Optimization Forecast Combination Framework

The problem is approached as that of finding the forecasts that are as close as possible to the preliminary figures that satisfy the required restrictions. In particular we focus on a least-squares squares formulation that translates into letting the undefined criterion for as close as possible to be governed by some quadratic loss function. We concentrate on solving the problem for one level of disaggregation, that is an aggregate and its components, as it is a setting that is relevant for many practical applications.

2.1.1 Formulating the Problem

The problem may be expressed as a general constrained quadratic program of the form:

$$\min_{\alpha,\beta} \sum_{i=1}^{A} f_{i,t}\left(y_{i,t},\alpha_{i,t},\varphi_{i,t}\right)^2 + \sum_{d=1}^{D}\sum_{n=1}^{N} g_{d,n,t}\left(q_{d,n,t},\beta_{d,n,t},\phi_{d,n,t}\right)^2 \tag{1}$$

subject to:

$$\left(1+\alpha_{1,t}\right)y_{1,t} - \sum_{n=1}^{N}\left(1+\beta_{1,n,t}\right)w_{1,n,t}q_{1,n,t} = 0$$

$$\left(1+\alpha_{1,t}\right)y_{1,t} - \left(1+\alpha_{i,t}\right)y_{i,t} = 0 \qquad \text{for } i = 2 \text{ to } A$$

$$\left(1+\beta_{1,n,t}\right)q_{1,n,t} - \left(1+\beta_{d,n,t}\right)q_{d,n,t} = 0 \qquad \text{for } d = 2 \text{ to } D,\ n = 1 \text{ to } N$$

where $y_{i,t}$ is the preliminary forecast for time t of the i-th aggregate model of a total of A, $\alpha_{i,t}$ is the percentage deviation of the definitive forecast from the preliminary, $\varphi_{i,t}$ is its exogenously chosen optimization weight and $f_{i,t}$ is some function of the three. Similarly, $q_{d,n,t}$ is the preliminary forecast for time t for component n of the d-th model of a total of D disaggregate models, $\beta_{d,n,t}$ is the percentage deviation of the definitive forecast from the preliminary, $\phi_{d,n,t}$ is its exogenously chosen optimization weight, $g_{d,n,t}$ is some function of the three and $w_{d,n,t}$ is the respective aggregation weight.[5]

The accounting identities are reflected directly in the constraints, but determining an appropriate loss function for the minimization problem is not straightforward. The literature on forecast combination is of little help because it has not dealt with the issue of combining different levels of aggregation in this way.[6] The reconciliation literature on the other hand has several suggestions, but given that they have been developed for a

[5]It is worth mentioning that all variables are in levels and that for simplicity it is assumed that all components and aggregation weights are strictly positive.

[6]Hyndman et al. (2011) formulate an unconstrained problem on the aggregation matrix.

different purpose it is necessary to make sure that they are adequate for the combination context.

We proceed by finding a loss function that in a setting where using a traditional single-variable forecasting combination method is feasible it produces the same outcome. In particular, we concentrate on the equal-weighted average due to its robust performance.

The following two assumptions provide the foundations for a setting where this method may be used:

1. The reliability of all forecasts are known to be the same.
2. All the information relevant for forecasting contained in the components is transmitted to the aggregate level.

These assumptions make working with the components equivalent to only dealing with their sum. In this context, the solution for this basic setting, using the nomenclature of equation (1), is:

$$\tilde{y}_t = \frac{1}{A+D}\left(\sum_{i=1}^{A} y_{i,t} + \sum_{d=1}^{D}\sum_{n=1}^{N} w_{d,n,t} q_{d,n,t}\right) \tag{2}$$

Although this setting could be seen as unrealistic, simple combination schemes are used extensively and there is ample evidence that in practice they often perform better than more involved procedures (Timmermann, 2006). In fact, the relative performance and robustness of the equal-weighted forecast combination is such that it has raised interest among researcher to try to explain it and has come to be known as the forecast combination puzzle (Smith and Wallis, 2009).

2.1.2 A Joint Combination Method for a Single Set of Forecasts

In developing a method that combines aggregate and disaggregate forecasts we start by focusing only on one set of forecasts. That is, for some period t, for a composite index X that is constructed by summing N components x_n using the respective time-varying aggregation weights w_n, there is a direct forecast y_t and a set of forecasts for its components $q_{n,t}$. In this context, two popular approaches that come from the reconciliation literature are the proportional and additive distribution methods. The proportional approach penalizes percentage deviations from the preliminary forecasts which translates into the loss function:

$$\varphi_t \left[\frac{(1+\alpha_t)y_t - y_t}{y_t}\right]^2 + \sum_{n=1}^{N} \phi_{n,t}\left[\frac{(1+\beta_{n,t})q_{n,t} - q_{n,t}}{q_{n,t}}\right]^2 \quad = \quad \varphi_t \alpha_t^2 + \sum_{n=1}^{N} \phi_{n,t}\beta_{n,t}^2 \tag{3}$$

Assigning discrepancies proportionally means bigger components absorb a larger share of the total adjustment. On the contrary, the additive approach attempts to evenly spread out the discrepancies among the variables. The associated loss function is:

$$\begin{aligned} &\varphi_t \left[(1+\alpha_t)\, y_t - y_t\right]^2 + \textstyle\sum_{n=1}^{N} \phi_{n,t} \left[(1+\beta_{n,t})\, q_{n,t} - q_{n,t}\right]^2 \\ &= \varphi_t \left(\alpha_t y_t\right)^2 + \textstyle\sum_{n=1}^{N} \phi_{n,t} \left(\beta_{n,t} q_{n,t}\right)^2 \end{aligned} \tag{4}$$

Unfortunately, both approaches applied directly fail to arrive at the desired outcome. On the one hand, the solution from a proportional approach is invariable strictly lower than the simple average of the aggregate forecasts. On the other hand, once aggregate and components are included into the same problem the outcome from the additive approach presents a bias towards the preliminary aggregate forecast. Although neither of the approaches produce the desired results when applied directly, it is possible to develop an appropriate loss function by recovering their respective desirable features.[7]

We start from equation (4) but impose a larger penalty term on the components so as to eliminate the aforementioned bias. Doing this results in the loss function being:

$$\varphi_t \left(\alpha_t y_t\right)^2 + Q_t \sum_{n=1}^{N} \phi_{n,t} w_{n,t} q_{n,t} \beta_{n,t}^2 \tag{5}$$

with $Q_t = \sum_{n=1}^{N} \left(w_{n,t} q_{n,t}\right)$.

Using this loss function and minimizing it subject to $(1+\alpha_t) y_t - \sum_{n=1}^{N} w_{n,t}(1+\beta_{n,t}) q_{n,t}$ gives as a solution that the definitive aggregate forecast for X_t is:[8]

$$\tilde{y}_t = \tilde{Q}_t = \frac{Q_t^2 + y_t \displaystyle\sum_{n=1}^{N} \left(\frac{\varphi_t}{\phi_{n,t}} w_{n,t} q_{n,t}\right)}{Q_t + \displaystyle\sum_{n=1}^{N} \left(\frac{\varphi_t}{\phi_{n,t}} w_{n,t} q_{n,t}\right)} \tag{6}$$

and the definitive forecast for the component $x_{n,t}$ is:

$$\tilde{q}_{n,t} = \left(1 + \frac{\varphi_t}{\phi_{n,t}} \cdot \frac{y_t - Q_t}{Q_t + \sum_{n=1}^{N} \left(\frac{\varphi_t}{\phi_{n,t}} w_{n,t} q_{n,t}\right)}\right) q_{n,t} \tag{7}$$

For the case of equal reliability, that is making $\varphi_{i,t}$ and $\phi_{d,n,t}$ equal to one, it becomes

[7] All this is shown in detail in section A.1 of the Appendix.

[8] This is shown in detail in section A.1.2 of the Appendix.

clear that equation (6) becomes a simple average. That is:

$$\tilde{y}_t = \frac{Q_t^2 + y_t Q_t}{2Q_t} = \frac{Q_t + y_t}{2} \tag{8}$$

2.1.3 A Joint Combination Method for Multiple Sets of Forecasts

For one set of forecasts the loss function suggested in the previous section results in the desired outcome. If more than one set of forecasts is considered for each variable however the outcome of the equal reliability case is not equal to the simple average.[9] Fortunately the bias that appears can be avoided simply by combining the multiple forecasts for the individual series before performing the joint combination and choosing the optimization weights so as to reflect the prior step.

Let the result for the prior step be:

$$y_t = \frac{1}{\Gamma_t}\sum_{i=1}^{A}\gamma_{i,t}y_{i,t} \qquad \text{and} \qquad q_{n,t} = \frac{1}{\Delta_{n,t}}\sum_{d=1}^{D}\delta_{d,n,t}q_{d,n,t} \tag{9}$$

with $\gamma_{i,t}$ and $\delta_{d,n,t}$ being the reliability weights, $\Gamma_t = \sum_{i=1}^{A}\gamma_{i,t}$ and $\Delta_{n,t} = \sum_{d=1}^{D}\delta_{d,n,t}$.

The joint combination procedure remains unchanged except for the weights φ_t and $\phi_{n,t}$ that are set to reflect the reliability of the combined forecasts y_t and $q_{n,t}$ as opposed to the initial preliminary forecasts $y_{i,t}$ and $q_{d,n,t}$.

In the case of equal reliability, this means accounting for the fact that the problem as a whole involves A aggregate and D disaggregate forecasts. That is accomplished by setting $\varphi_t = A$ and $\phi_{n,t} = D$ making the solution for the aggregate forecast:

$$\tilde{y}_t = \frac{1}{A+D}\left(A \cdot y_t + D \cdot \sum_{n=1}^{N} w_{n,t}q_{n,t}\right) \tag{10}$$

By expanding the individual forecasts, given that $\gamma_{i,t}$ and $\delta_{d,n,t}$ are equal to one, the definitive aggregate forecast is left in terms of the preliminary estimates:

$$\begin{aligned}\tilde{y}_t &= \tfrac{1}{A+D}\left(A \cdot \tfrac{1}{A}\textstyle\sum_{i=1}^{A} y_{i,t} + D \cdot \sum_{n=1}^{N} \tfrac{1}{D} w_{n,t} \sum_{d=1}^{D} q_{d,n,t}\right) \\ &= \tfrac{1}{A+D}\left(\textstyle\sum_{i=1}^{A} y_{i,t} + \sum_{d=1}^{D}\sum_{n=1}^{N} w_{n,t}q_{d,n,t}\right)\end{aligned} \tag{11}$$

that is the same as taking the simple average of all the available forecasts for the aggregate.

[9] This is shown in section A.1.3 of the Appendix,

This result shows that the method replicates the outcome of the equal-weighted forecast combination for the aggregate while at the same time providing component forecasts that are fully consistent. More generally however, the framework admits taking into consideration the reliability of the different forecasts.[10] This is a desirable feature when the uncertainty surrounding the different forecast differs like, for example, in the case of nowcasting where necessary inputs may include both preliminary and definitive figures.

2.1.4 Feasible Region for the Reliability Weights

From the solution in equation (6) we notice that what matters is the relative reliability and that therefore any given number in isolation is meaningless. It is important however to establish a feasible region that provides a unique solution for the minimization problem. We do this by looking at the bounds for the weights and what they imply regarding overall reliability. Considering as a starting point that all weights are set equal to one:

1. *Absolute Certainty*: The limit for a high degree of reliability is to eliminate all uncertainty from the outcome. For the aggregate forecast this means making φ_t go to infinity. In such a case it is easy to see that $\lim_{\varphi_t \to \infty} (1 + \alpha_t)\, y_t = y_t$. On the other hand, for a single component $n = 1$, setting φ_t back to one and making $\phi_{1,t}$ go to infinity implies that $\frac{\varphi_t}{\phi_{1,t}} \to 0$. This means that the weight given to the direct forecast decreases but still remains positive. Taking it to the extreme and making all component's weights go to infinity decreases the weight given to the direct forecast to zero. That is $\lim_{\phi_t \to \infty} (1 + \alpha_t)\, y_t = Q_t$ where $\phi_{n,t} = \phi_t$ for $n = 1$ to N. All forecasts however cannot be certain otherwise the problem does not have a solution. This means the ceiling for reliability weights is infinity but at least one of them has to be finite.

2. *Zero Confidence*: The opposite to a high degree of reliability is to have absolutely no confidence whatsoever in a forecast. For the aggregate forecast this would mean making $\varphi_t = 0$ and therefore $\tilde{y}_t = Q_t$. On the converse, for a single component $n = 1$, setting φ_t back to one and making $\phi_{1,t} = 0$ means that this component absorbs all the deviation. This is clear from appreciating that $\frac{\varphi_t}{\phi_{1,t}} \to \infty$ and therefore that $\lim_{\phi_{1,t} \to 0} (1 + \alpha_t)\, y_t = y_t$. This basically means that the forecasts from all but this component are taken as given and that the definitive forecast $\tilde{q}_{1,t}$ is found residually. It is worth noting that this can be done for one variable only otherwise the minimization problem has infinite solutions. This means that no more than one variable can have a reliability weight equal to zero for the problem to have a unique solution.

[10] For this case in which only one level of disaggregation is considered it is easy to show that the method proposed by Hyndman et al. (2011) is a particular case. This is shown in Appendix A.2.3

For the purpose of allowing for some degree of combination it makes sense to restrict the aggregate forecasts to have finite reliability weights. This means that a given component could have a weight that implies certainty, maybe due to the early release of relevant data, but not all of them.

3 Combining GDP Forecasts for Three European Economies

As an empirical application of the method we perform a forecasting exercise using GDP data from France, Germany and the United Kingdom. We use eight different forecasting models and four different ways of establishing the combination weights within our framework. We evaluate the aggregate forecasting accuracy by comparing the results with that of the single models and traditional forecast combinations. The forecasting accuracy of the components is evaluated against that of the single-models.

3.1 Data

For the exercise we use GDP series from both the production and expenditure approaches for France, Germany and the United Kingdom. The data is quarterly and seasonally adjusted, spanning from 1991 to 2014 and available from the OECD statistics database.[11]

As in most of the OECD, these countries calculate their GDP using a chain-linking method.[12] A well known rather unpleasant feature of this method is that the volume of changes in inventories cannot be constructed as chain-linked series (Lequiller and Blades, 2014). This is problematic because they are required in order to complete the aggregate GDP forecasts from the expenditure approach.

Faced with this problem, some authors proceed by expressing the series in terms of contributions to GDP growth and adjust their methods accordingly. In order to avoid dealing with a series that becomes close to zero and changes sign often we bundle change in inventories with imports. The rational behind that is that in practice changes in inventories to a great extent serve as a buffer for foreign trade and considering them together could be beneficial. This is supported by Esteves (2013) who find that forecasting errors

[11] For the United Kingdom the production data on the OECD database starts in 1995. The first four years of the sample are obtained by splicing backwards the historical reference tables available from the Office for National Statistics. No inconsistencies arise from the seasonal adjustment given that the aggregates are adjusted indirectly, that is as the sum of the seasonally adjusted components.

[12] According to the updated survey in OECD (2009) out of the members only Mexico uses a fixed-base method.

of imports and changes in inventories are highly correlated and that forecasting them jointly increases forecasting accuracy.

Taking all that into consideration, the breakdown of aggregate GDP for all three countries is the following:

Table 1: GDP Production and Expenditure Components

Production:		
	1. Agriculture, forestry and fishing	7. Financial and insurance activities
	2. Manufacturing	8. Real estate activities
	3. Industry and energy, excluding manufacturing	9. Professional, administrative and support service activities
	4. Construction	10. Public adm., defence, social security, education and health
	5. Trade, transport, accommodation and food services	11. Other service activities
	6. Information and communication	12. Taxes less subsidies
Expenditure:		
	1. Private consumption	4. Exports of goods and services
	2. Government expenditure	5. Imports of goods and services and changes in inv.
	3. Gross fixed capital formation	

3.2 Forecasting Models

Regardless of the numerous developments in econometric modelling, univariate methods continue to provide an often strong benchmark against which to compare other models (Marcellino, 2008). They are also the methods used in many of the aggregate-disaggregate forecasting competitions mentioned in the literature review and are therefore a reasonable starting point.

For this purpose we use a random walk for the growth rate, an autoregressive model of order one for the first differences of the variables and ARIMA models chosen following a common and well established routine. In particular we rely on the program TRAMO (Gomez and Maravall, 1996) that through an automatic procedure selects the appropriate transformation and chooses the model based on the Bayesian Information Criterion (BIC).

To account for the interdependence between components we also use Bayesian Vector Autoregressive models (BVARs) following the implementation in Banbura et al. (2010). In particular, the two first sets of VARs include the Consumer Price Index (CPI) and the respective approach, that is: only the aggregate GDP, only the production side components and only the expenditure components, estimated with all variables in first differences and also differentiating CPI twice.

Following the notion in Hendry and Hubrich (2011) we also estimate VARs that include all the GDP variables and CPI in the same model. We cast the VARs in levels, first differences and in first differences with CPI differentiated twice.

The smallest VARs, that is the two that include CPI and only the aggregate GDP, are estimated by OLS using two lags. All the others are estimated using five lags and the choice of overall tightness, as in Banbura et al. (2010), is made such that the in-sample fit equals that of a two-variable VAR with five lags estimated by OLS over the first 10 years of the sample.

All this results in eight sets of forecasts over the forecasting horizon for each one of the variables.

3.3 Forecasting Accuracy Comparison

3.3.1 Set-up of the Evaluation Exercise

The evaluation exercise is performed over the 2001-2014 period leaving the first years of data to estimate the models. It is set up in a quarterly rolling scheme using a ten year window where in each period the models are re-estimated and a one-year-ahead quarterly forecast is generated.

The forecasting accuracy is presented, for different horizons, by means of the model's mean square forecasting error (MSFE) relative to that of a benchmark model. That is, for variable i, horizon h and using model m, the relative MSFE is

$$\mathrm{RelMSFE}^{(i,h,m)} = \frac{\mathrm{MSFE}_{T_0,T_1}^{(i,h,m)}}{\mathrm{MSFE}_{T_0,T_1}^{(i,h,0)}}$$

with

$$\mathrm{MSFE}_{T_0,T_1}^{(i,h,m)} = \frac{1}{T_1 - T_0 + 1} \sum_{t=T_0}^{T_1} \left(y_{i,t+h}^{(m)} | t - y_{i,t+h} \right)^2$$

where $y_{i,t+h}^{(m)}|t$ is the forecasted value for $t+h$ at time t and T_0 is the last period of actual data in the first sample used for the evaluation and T_1 is the last period of actual data in the last sample. As usual a RelMSFE lower than one reflects an improvement over the benchmark model for which $m = 0$.

Regarding measuring the overall forecasting accuracy of the components we do so by comparing the cumulative absolute errors in the contribution to the aggregate level. For this purpose we define the cumulative absolute root mean square forecasting error for an aggregate with N components q_n, horizon h and using model m as

$$\mathrm{CumRMSFE}_{T_0,T_1}^{(h,m)} = \sqrt{\frac{1}{T_1 - T_0 + 1} \sum_{t=T_0}^{T_1} \left(\sum_{n=1}^{N} w_{n,t+h} \cdot \text{abs} \left(q_{n,t+h}^{(m)} | t - q_{n,t+h} \right) \right)^2}$$

where $q^{(m)}_{n,t+h}|t$ is the forecasted value for $t+h$ at time t and T_0 is the last period of actual data in the first sample used for the evaluation and T_1 is the last period of actual data in the last sample.

3.3.2 Specific Forecast Combination Minimization Problem

The empirical exercise contemplates combining forecasts for GDP from direct approaches and disaggregate approaches from the production and expenditure sides.

Let the result of the prior combination step be a unique direct forecast y, a production side forecast Q based on the N components q_n and an expenditure side forecast G based on the M components g_m. Then, the minimization problem involving the aggregate reliability weight φ, the production reliability weights ϕ_n, the expenditure reliability weights ψ_m, the production aggregation weights w_n and the expenditure aggregation weights ω_m, is:

$$\min_{\alpha,\beta,\epsilon} \varphi_t \left(\alpha_t y_t\right)^2 + Q_t \sum_{n=1}^{N} \phi_{n,t} w_{n,t} q_{n,t} \left(\beta_{n,t}\right)^2 + G_t \sum_{m=1}^{M} \psi_{m,t} \omega_{m,t} g_{m,t} \left(\epsilon_{m,t}\right)^2 \tag{12}$$

subject to:

$(1+\alpha_t)y_t - \sum_{n=1}^{N} w_{n,t}(1+\beta_{n,t})q_{n,t} = 0$

$(1+\alpha_t)y_t - \sum_{m=1}^{M} \omega_{m,t}(1+\epsilon_{m,t})g_{m,t} = 0$

Similarly to the simple setting, this problem also has analytical solutions.[13] The definitive aggregate forecast also follows a weighted average form:

$$\tilde{y}_t = \frac{\Upsilon_t \cdot y_t + \Omega_t \cdot Q_t + \Xi_t \cdot G_t}{\Upsilon_t + \Omega_t + \Xi_t} \tag{13}$$

where $\Upsilon_t = \sum_{n=1}^{N} \left(\frac{\varphi_t}{\phi_{n,t}} w_{n,t} q_{n,t}\right) \cdot \sum_{m=1}^{M} \left(\frac{\varphi_t}{\psi_{m,t}} \omega_{m,t} g_{m,t}\right)$, $\Omega_t = Q_t \cdot \sum_{m=1}^{M} \frac{\varphi_t}{\psi_{m,t}} \omega_{m,t} g_{m,t}$ and $\Xi_t = G_t \cdot \sum_{n=1}^{N} \frac{\varphi_t}{\phi_{n,t}} w_{n,t} q_{n,t}$. The definitive forecasts for all of the components can in turn then be easily recovered using a solution similar to that of equation (7).

3.3.3 Establishing Reliability Weights

Even in the absence of relevant external knowledge it may be desirable to determine reliability weights based on the properties of the preliminary estimates. Timmermann

[13] These are provided in section A.2.2 of the Appendix,

(2006) present an extensive survey on some of the suggestions from the combination literature for single variables and more become available from ongoing research (Hansen, 2008; Wei and Yang, 2012; Hsiao and Wan, 2014).

Taking into consideration the easiness with which each suggestion can be incorporated into our framework we choose a few.

Equal Weights:

An obvious choice for the first set of weights is equal weights because it serves as the benchmark against which to compare all the others.

In-Sample Fit:

Using in-sample fit to determine combination weights is not uncommon. Kapetanios et al. (2008) find promising results from using weights calculated using information criteria. Extending this approach to compare different series is not straightforward. We can however extend the way that Banbura et al. (2010) determine in-sample fit for their Bayesian VARs by normalizing the measure.

We start by defining the root mean square percentage error (RMSPE) at time u using information up to time p for the h-step ahead forecast of x_i as:

$$RMSPE_{i,u,p,h,v} = \sqrt{\frac{1}{v}\sum_{s=u-h-v}^{u-h}\left(\frac{x_{i,s+h}|p}{x_{i,s+h}} - 1\right)^2} \tag{14}$$

where $x_{i,s+h}|p$ is the fitted value for x_i using the coefficients calculated at time p and v determines how much data is included in the measure. The latter is limited by the number of lags that are included in each model.

We then define the weights based on in-sample fit as:[14]

$$\omega_{i,t,h,v}^{ISP} = \frac{1}{RMSPE_{i,t,t,h,v}} \tag{15}$$

Out-of-Sample Past Performance:

An obvious extension of the idea of weighting according to predictability is to weigh the different forecasts based on their recent past relative out-of-sample performance. This approach goes as far back as Bates and Granger (1969). Empirical studies suggest that

[14]In the context of forecast combination using predictive measures, Eklund and Karlsson (2007) raise awareness regarding the possibility of distorting weight distribution due to overconfidence in models that over-fit data. Aiolfi and Favero (2005) for example use the model's R^2 to decide on the combination of forecasts.

forecasts weighted by the inverse of their MSE are found to work well in practice (Stock and Watson, 1999; Timmermann, 2006).

Following the same idea and arguments expressed for the in-sample fit weights, we define the weights based on out-of-sample past performance as:

$$\omega_{i,t,h,v}^{OSP} = \frac{1}{RMSPE_{i,t,s,h,v}} \tag{16}$$

where in this case the s that goes into the formula as the time subscript is not a parameter, but the index in the sum embedded in equation (14).

Optimal weights:

In the context of single variable combinations Granger and Ramanathan (1984) address the problem of determining the optimal combination weights as a least-squares regression problem. Hyndman et al. (2011) extend the approach to a setting with variables from different aggregation levels. In their implementation however they do not consider forecasts from more than one hierarchical order. To enable combining forecasts from both the production and expenditure approaches we use an approximation and set the weights to:[15]

$$\omega_{i,t}^{\text{OPT}} = \frac{x_{i,t}}{\sum_{n=1}^{N} x_{i,t}} \tag{17}$$

In the case of all weighting methods the reliability weights are calculated for every rolling window. For the in-sample the five most recent years of the window are used and for the out-of-sample weights the last two.

3.3.4 Aggregation weights

Before being able to perform the forecasting exercise one final point needs to be addressed. Given the chain-linked nature of the series, the aggregation weights required for the forecasting process are not fixed. This is a relevant issue as Lütkepohl (2011) and Brüggemann and Lütkepohl (2013) show that accounting for the changing weights can increase forecasting accuracy significantly.

In the chain-linking framework aggregation weights depend on previous year prices and therefore for short forecasting horizons some information regarding the prices that will go into the weights is available. To take this issue into consideration without having to actually forecast the aggregation weights, we use a simple adaptive procedure to

[15]The derivation shown in section A.2.3 of the Appendix

Table 2: Single Model Aggregate Relative Forecasting Errors by Approach

	Direct				Production				Expenditure			
Horizon	1	2	3	4	1	2	3	4	1	2	3	4
France												
MIN	0.84	0.83	0.85	0.86	0.87	0.90	0.94	0.97	0.88	0.90	0.92	0.93
MAX	1.00	1.00	1.03	1.10	1.00	1.00	1.03	1.08	1.00	1.00	1.03	1.14
MEDIAN	0.90	0.93	0.99	1.01	0.93	0.96	1.00	1.03	0.93	0.95	0.99	1.00
Germany												
MIN	0.94	0.96	0.97	0.99	0.93	0.99	1.01	1.01	0.97	1.01	1.02	1.02
MAX	1.02	1.09	1.13	1.18	1.03	1.12	1.13	1.14	1.06	1.28	1.37	1.40
MEDIAN	1.00	1.05	1.09	1.12	1.00	1.04	1.06	1.09	1.00	1.04	1.07	1.09
UK												
MIN	0.71	0.78	0.87	0.91	0.71	0.81	0.89	0.93	0.71	0.81	0.88	0.92
MAX	1.00	1.00	1.02	1.11	1.01	1.01	1.01	1.02	1.13	1.11	1.11	1.10
MEDIAN	0.73	0.83	0.92	0.96	0.83	0.90	0.95	0.97	0.84	0.90	0.94	0.96

Note: Minimum, median and maximum of the mean square forecasting error of the individual models relative to that of the direct approach using the random walk model for each horizon. The individual models are a random walk with drift, a first-differences autoregressive model of order one, an ARIMA chosen according to the Bayesian Information Criterion, two small VARs including CPI and the GDP variables from each approach in first differences and where CPI is differenced twice and three large VARs including CPI and all GDP variables in levels, in first differences and in first differences with CPI differenced twice. Calculated for one to four steps ahead forecasts over the 2001-2014 period.

provide estimates of the unavailable weights. This procedure consists on using the implicit deflator that results from the most recent four-quarter moving averages of the non-seasonally adjusted nominal and chain-linked series to have an updated estimate of the future aggregation weights.[16]

3.4 Results

3.4.1 Forecast Combination Accuracy Over the Whole Sample

The forecasting application involves eight different forecasting models and three different approaches. Table 2 presents the minimum, median and maximum of the individual models' relative forecasting accuracy for the aggregate over the 2001-2014 sample for the three countries.[17]

Looking at the differences between countries it becomes apparent that for France and the UK there are quite significant improvements over the naive random walk model while in the case of Germany most models do worse. For all three countries however the direct approach tends to achieve the best results.

[16] According to the survey in OECD (2009) updated to April 2015, almost 72% of the OECD members and the OECD itself use the annual overlap method for their quarterly accounts and therefore we use this method for this exercise. It should be noted that the United Kingdom uses the slightly different quarterly overlap method.

[17] The relative forecasting accuracy of each model is provided in section B.1.1

Table 3: Combination Aggregate Relative Forecasting Error

Horizon	Aggregate 1	2	3	4	Joint 1	2	3	4
France								
Eq.W.	0.86	0.88	0.92	0.94	0.86	0.88	0.92	0.94
ISP	0.87	0.89	0.93	0.96	0.87	0.89	0.93	0.96
OSP	0.86	0.88	0.92	0.95	0.86	0.88	0.91	0.94
OPT	0.95**	0.98**	1.02**	1.02**	0.86	0.87	0.90	0.93
Germany								
Eq.W.	0.95	1.00	1.04	1.06	0.95	1.00	1.04	1.06
ISP	0.96	1.01	1.04	1.07	0.96	1.01	1.04	1.07
OSP	0.96	1.01	1.04	1.07	0.95	1.00	1.04	1.07
OPT	1.02	1.11	1.17	1.21	0.95	1.00	1.03	1.05
UK								
Eq.W.	0.76	0.83	0.88	0.91	0.76	0.83	0.88	0.91
ISP	0.74	0.83	0.89	0.92	0.73	0.82	0.88	0.92
OSP	0.75	0.83	0.89	0.92	0.74	0.82	0.88	0.91
OPT	0.73	0.88	1.00	1.04	0.74	0.81	0.88	0.91

Note: Mean square forecasting error of each model relative to that of the direct approach using the random walk model for each horizon. The combination weighting schemes are the simple average (EQ.W), in-sample fit (ISP), out-of-sample performance (OSP) and optimal weights (OPT). For the aggregate optimal weights we use the approach in Conflitti et al. (2015) that impose the constraints that weights should be non-negative and sum up to one. * and ** denote that the respective forecast is statistically worse than the best single model within the sample according to the Modified Diebold-Mariano statistic at a 10 and 5% significance level. Calculated over the 2001-2014 period.

The overall dispersion in forecasting accuracy between single models varies among countries.[18] For France the comparison between the minimums achieved for each horizon with the maximums show that the worst performing models are between 20 and 30% less accurate depending on the horizon while the same comparison for the median show differences of 10 to 20%. For Germany the same comparison shows that the worst performing models are between 10 and 20% less accurate while for the median differences are around of 10%. For the UK the differences are between 20 and 40% and around 10% respectively.

Given the varying performance of the models it could turn out to be quite costly choosing one only. The appeal of forecast combination is that this is not necessary and Table 3 presents their RelMSFE for this exercise. In the column under the “Aggregate” heading we present the outcome for the traditional forecast combination for single variables applied to the aggregates that result from all three approaches calculated using equivalent weighting schemes.[19] In the column under the “Joint” heading we present the outcome of our framework.

The first thing to notice is that overall the differences between weighting schemes are very small within each approach. The only exception is the aggregate optimal combination. The second is that there is hardly no difference between the aggregate accuracy

[18] The outstandingly bad ARIMA for expenditure approach is removed for the analysis of maximums.

[19] That is equal-weights, in-sample fit, out-of-sample performance and optimal weights as in Conflitti et al. (2015).

Table 4: Cumulative Disaggregate Relative Forecasting Error

	Production				Expenditure			
Horizon	1	2	3	4	1	2	3	4
France								
Single Model Median	1.06	1.07	1.06	1.05	1.06	1.08	1.09	1.09
Combination								
Eq.W.	1.00	0.99	0.98	0.97	0.99	1.00	1.01	1.02
ISP	1.00	1.00	0.99	0.99	1.00	1.02	1.03	1.04
OSP	0.99	0.98	0.98	0.97	0.99	1.01	1.01	1.02
OPT	1.00	0.99	0.98	0.97	0.98	1.00	1.00	1.01
Germany								
Single Model Median	1.04	1.02	1.02	1.03	1.02	1.04	1.06	1.06
Combination								
Eq.W.	1.01	0.99	1.00	1.01	0.97	0.99	1.01	1.01
ISP	1.01	0.99	1.01	1.01	0.98	0.99	1.00	1.01
OSP	1.01	0.99	1.00	1.01	0.98	0.99	1.00	1.00
OPT	1.01	0.99	1.00	1.01	0.98	0.99	1.01	1.01
UK								
Single Model Median	1.04	1.03	1.02	1.02	1.08	1.12	1.12	1.11
Combination								
Eq.W.	1.00	0.99	0.98	0.98	1.02	1.04	1.04	1.03
ISP	0.99	0.99	0.98	0.98	1.02	1.02	1.02	1.02
OSP	0.99	0.99	0.98	0.98	1.03	1.04	1.03	1.03
OPT	0.99	0.98	0.97	0.97	1.02	1.04	1.04	1.03

Note: Cumulative root mean square forecasting error of each combination method relative to the minimum achievable from the single models for each horizon. The combination weighting schemes are the simple average (EQ.W), in-sample fit (ISP), out-of-sample performance (OSP) and optimal weights (OPT). Calculated over the 2001-2014 period.

of the joint combination and its traditional counterpart. This means that no harm in this sense is caused by using our framework. In fact, for the UK the joint combination performs marginally better for most weighting schemes. Regarding the actual performance, as one would expect given that all models enter with positive weights, the minimum RelMSFE from the single models are not achieved. They do however come well below the median of the single models.

In regards to the disaggregate accuracy, Table 4 presents the CumRMSFE of the joint method for both the production and expenditure approaches relative to that of the best single model within each approach for each horizon. The results show that some of the features present at the aggregate level translate to the components. In particular the overall differences between weighting schemes are very small within each approach.

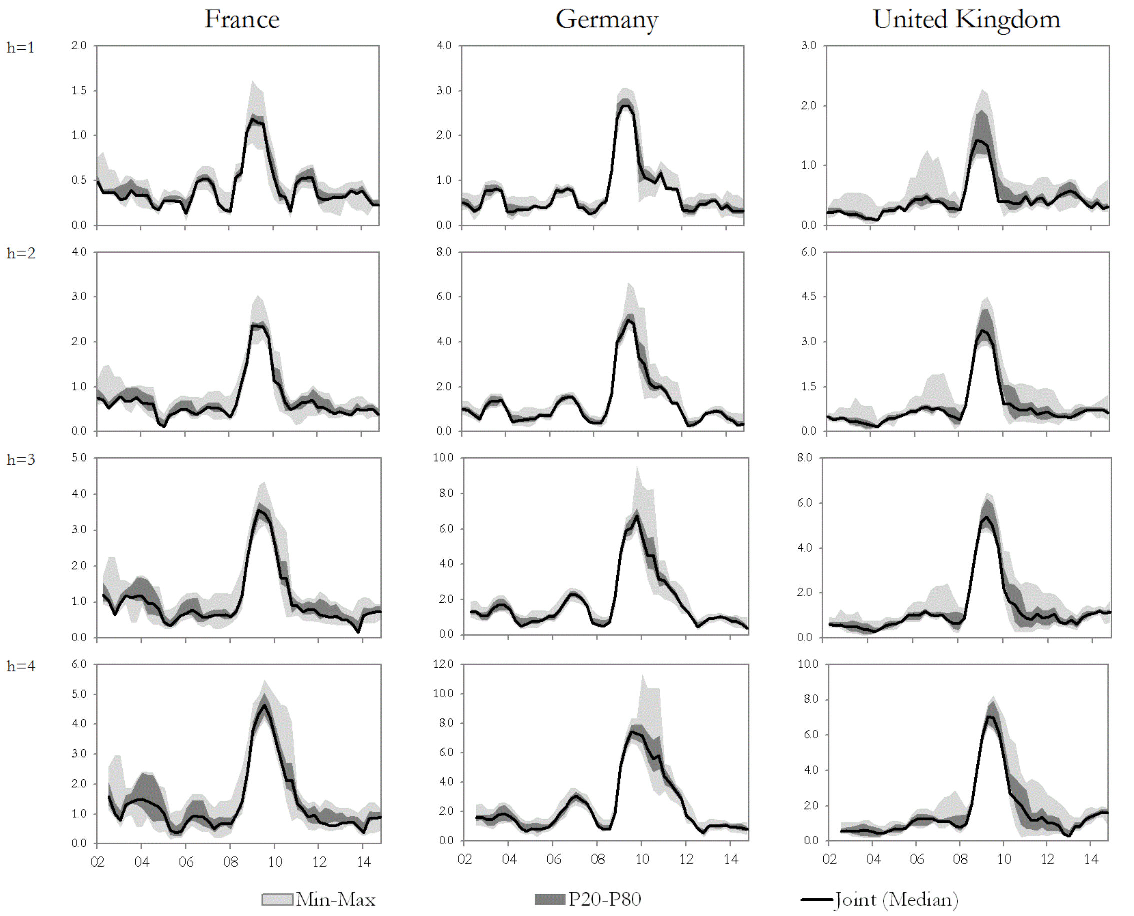
The most remarkable result however is the fact that for all three countries the joint combination often improves on the best performing single model. In the case of France for the production approach the accuracy from most of the combination methods is at least as good as the best single model and improves up to 3%. For the expenditure approach on the other hand it improves up to 2% but mostly equals or is up to 2% worse. With regards to the median the combinations are approximately 7% better for both production and expenditure. In the case of Germany the major improvements are found for the expenditure approach with up to 3%. Overall for both approaches the accuracy of the combination are very similar to the best model and from 2 to 5% better than the median. For the UK there is an overall improvement for the production approach of up to 3%. The expenditure approach on the other hand does no improve but remains quite close to the best model. Overall for both approaches the accuracy of the combination is from 4 to 8% better than the median.

Overall we find that the joint combination methods perform well given that most of them achieve similar accuracy as the best performing single model of each approach in a context where the best single models of each approach are not necessarily consistent with each other or as good as the combined forecast in terms of aggregate accuracy.

3.4.2 Aggregate Forecasting Accuracy Over the Evaluation Sample

One non-trivial detail of our forecasting exercise is that the evaluation period includes the end portion of what has been called the Great Moderation and the 2008 world financial crisis. A considerable body of literature has devoted itself to understand the effects of these periods on forecasting models and Chauvet and Potter (2013) present a comprehensive review. Some of the conclusions include that many well performing models in

Figure 1: Dispersion of the Rolling Forecasting Error



Note: Four-quarter rolling root mean square forecasting error (RMSFE) for each horizon. The Min-Max shaded area shows the span between the minimum and maximum RMSFE from the 24 individual models/approaches in each period. The P20-P80 does the same but trims off the top and bottom 20%. joint is the median of the joint combination methods. Calculated as four-quarter moving windows over the 2001-2014 period.

stable times completely failed with the increase of volatility and that models perform differently in expansions and recessions. This last point had been previously documented in Marcellino (2008) who find that in recessions their more sophisticated models showed a marked deterioration making the simple random walk the best performer.

For the purpose of our exercise this could lead to our results being overly influenced by the particular performance in the crisis years simply because the forecasting errors could be massive. It makes sense therefore to look at how the forecasting errors evolve over the sample. To do so, we look at the four-quarter rolling root mean square error for all forecasting horizons. Figure 1 presents the dispersion of the single models referred to as min-max, the same measure trimming the best and worst performing 20% and the median of the joint combinations.

The impact of the financial crisis is obvious for all countries and over all horizons and, as suspected, the size of the forecasting errors would make this period predominant in the overall results.

Regarding other aspects, the picture looks relatively similar across forecasting horizons within countries but a bit different for each country. For France the dispersion of the models seems relatively high before and after the financial crisis and not so much during it and the effect of the crisis on forecasting accuracy is relatively short-lived. For Germany on the other hand dispersion is relatively low over the whole sample but the effects of the crisis on accuracy go on for much longer than for the other two countries. For the UK the dispersion is relatively low before the crisis, but high during it and remains moderately high thereafter. The forecasting errors decrease really fast after the crisis.

Regarding the performance of the combination method, the median measure registers values at or very close to the lower boundary of the trimmed dispersion measure for most of the evaluation sample.

The way in which errors evolve over the sample suggest that the crises years could be too determinant in the overall results. We therefore perform the previous analysis excluding years 2008 and 2009 from it. The episode and its consequences are bound to be long lasting and for this reason, although we remove its direct impact on the measure for forecasting accuracy, the effects on the estimation of the parameters remain.

As before, Table 5 presents the relative forecasting accuracy for the aggregate of the single models but this time for the restricted sample.

The changes are quite dramatic. For France and the UK the improvements of the models over the random walk completely disappear. In fact, most of the models turn out to be significantly worse. Only for Germany does the general picture look similar. Also, in this case it is the bottom-up production side approach that shows marginally better results.

The significant increase in overall dispersion in forecasting accuracy between single models for both France and the UK is clear from comparing the minimum and the median. For the former it goes up to 10 to 30% depending on the horizon and the latter to 10 to 40%. For Germany on the other hand the same measurement remains around 10%.

The performance of the forecast combination however does not appear that different as it can be seen from Table 6. Differences between weighting schemes remain very small within each approach with the exception of the optimal combination. Again there is hardly no difference between the aggregate accuracy of the joint combination and

Table 5: Single Model Forecasting Errors excluding 2008-2009

Horizon	Direct 1	2	3	4	Production 1	2	3	4	Expenditure 1	2	3	4
France												
MIN	1.00	1.00	1.00	1.00	1.00	0.95	1.01	1.01	1.01	1.01	1.01	1.03
MAX	1.16	1.26	1.42	1.56	1.17	1.27	1.42	1.57	1.23	1.31	1.44	1.58
MEDIAN	1.06	1.06	1.13	1.17	1.13	1.22	1.35	1.45	1.13	1.18	1.25	1.31
Germany												
MIN	0.96	0.94	0.93	0.94	0.87	0.90	0.93	0.92	0.94	0.94	0.91	0.92
MAX	1.03	1.08	1.13	1.17	1.11	1.28	1.24	1.18	1.04	1.17	1.19	1.18
MEDIAN	1.00	0.99	1.01	1.01	1.02	1.03	1.04	1.06	0.99	1.00	1.01	1.00
UK												
MIN	1.00	1.00	1.00	1.00	1.05	0.97	0.98	0.98	1.05	1.12	1.04	1.04
MAX	1.18	1.29	1.42	1.55	1.18	1.29	1.41	1.53	1.89	2.12	2.12	2.20
MEDIAN	1.12	1.19	1.33	1.40	1.10	1.21	1.35	1.45	1.15	1.26	1.34	1.43

Note: Minimum, median and maximum of the mean square forecasting error of the individual models relative to that of the direct approach using the random walk model for each horizon. The individual models are a random walk with drift, a first-differences autoregressive model of order one, an ARIMA chosen according to the Bayesian Information Criterion, two small VARs including CPI and the GDP variables from each approach in first differences and where CPI is differenced twice and three large VARs including CPI and all GDP variables in levels, in first differences and in first differences with CPI differenced twice.. Calculated for one to four steps ahead forecasts over the 2001-2014 period excluding years 2008 and 2009.

Table 6: Combination Aggregate Forecasting Error excluding 2008-2009

Horizon	Aggregate 1	2	3	4	Joint 1	2	3	4
France								
Eq.W.	1.01	1.00	1.06	1.12	1.01	1.00	1.06	1.12
ISP	1.02	1.03	1.11	1.18*	1.02	1.03	1.10	1.17*
OSP	1.01	1.00	1.05	1.10	1.01	1.00	1.04	1.08
OPT	1.17*	1.26*	1.37**	1.46**	1.01	1.00	1.06	1.11
Germany								
Eq.W.	0.95	0.98	0.99	1.00	0.95	0.98	0.99	1.00
ISP	0.95	0.98	1.00	1.01	0.95	0.97	0.99	1.01
OSP	0.95	0.98	1.00	1.01	0.95	0.97	0.99	1.01
OPT	1.09*	1.07*	1.09*	1.09	0.95	0.97	0.98	0.99
UK								
Eq.W.	1.03	1.08	1.15**	1.20*	1.03	1.08	1.15**	1.20*
ISP	1.03	1.09	1.17**	1.21*	1.03	1.09	1.17**	1.21*
OSP	1.03	1.08	1.15**	1.19*	1.03	1.08	1.15**	1.19*
OPT	1.13	1.20*	1.33**	1.39**	1.04	1.08	1.15**	1.19**

Note: Mean square forecasting error of each model relative to that of the direct approach using the random walk model for each horizon. The combination weighting schemes are the simple average (EQ.W), volatility (VOL), in-sample fit (ISP), out-of-sample performance (OSP), simple average for the first stage and optimal weights for the second (OPT-1) and optimal weights for both stages (OPT-2). * and ** denote that the respective forecast is statistically worse than the best single model within the sample according to the Modified Diebold-Mariano statistic at a 10 and 5% significance level. Calculated over the 2001-2014 period excluding 2008 and 2009.

Table 7: Cumulative Disaggregate Forecasting Error excluding 2008-2009

	Production				Expenditure			
Horizon	1	2	3	4	1	2	3	4
France								
Single Model Median	1.09	1.13	1.13	1.15	1.12	1.14	1.22	1.25
Combination								
Eq.W.	1.00	1.01	0.99	1.00	1.02*	1.05	1.09	1.12**
ISP	1.01	1.03	1.01	1.03	1.03**	1.08*	1.14	1.17**
OSP	1.00	1.01	0.99	0.99	1.02*	1.05	1.09	1.11**
OPT	1.01	1.02	1.00	1.01	1.02*	1.05	1.09	1.11*
Germany								
Single Model Median	1.04	1.03	1.03	1.03	1.07	1.10	1.15	1.13
Combination								
Eq.W.	1.01	1.02	1.02	1.01	1.03	1.04	1.08	1.08
ISP	1.01	1.01	1.01**	1.01	1.03	1.05	1.10	1.11
OSP	1.01	1.01	1.01**	1.01	1.03	1.04	1.09	1.09
OPT	1.02	1.02	1.02	1.02	1.03	1.04	1.08	1.08
UK								
Single Model Median	1.03	1.06	1.07	1.07	1.03	1.03	1.03	1.06
Combination								
Eq.W.	1.00	1.02	1.02	1.02	1.01	1.00*	1.00	1.02
ISP	1.00	1.02	1.02	1.02	1.03	0.99*	0.98	1.00
OSP	1.00	1.02	1.02	1.02	1.03	1.00*	0.99	1.01
OPT	1.00	1.02	1.01	1.02	1.01	1.00*	1.00	1.03

Note: Cumulative root mean square forecasting error of each combination method relative to the minimum achievable from the single models for each horizon. The combination weighting schemes are the simple average (EQ.W), volatility (VOL), in-sample fit (ISP), out-of-sample performance (OSP), simple average for the first stage and optimal weights for the second (OPT-1) and optimal weights for both stages (OPT-2). * and ** denote that the respective forecast is statistically worse than the best single model within the sample according to the Modified Diebold-Mariano statistic at a 10 and 5% significance level. Calculated over the 2001-2014 period excluding years 2008 and 2009.

its traditional counterpart with some joint combination methods performing marginally better.

Regarding the actual performance however the RelMSFE of the combinations are further from that of the best single models. For France they are between 1 and 10% worse but still 15% better than the median model. For Germany they are about 9% worse and between 2 and 5% better than the median. For the UK they are between 5 and 20% worse and between 10 and 15% better than the median. Also, the performance of the combination methods deteriorates for the longer horizons.

In regards to the disaggregate accuracy, Table 7 presents the cumulative RMSFE of the joint method for both the production and expenditure approaches relative to that of the best single model within each approach for each horizon.

In this case the overall differences between weighting schemes are again very small within each approach. It is noteworthy however that, notwithstanding the increased dispersion in performance that results from removing the crisis from the sample, some improvements over the best performing single models are still found. This happens specifically for France on the production side and the UK on the expenditure side.

The overall performance looks quite good. In the case of France for the production approach the accuracy from most of the combination methods is very similar to that of the best single model. This is not achieved for the expenditure approach however that is up to 10% worse. Nevertheless, with regards to the median the combinations are approximately 10 to 15% better for both production and expenditure. Similarly, in the case of Germany the accuracy is similar for the production approach but suffers for the expenditure approach and exhibits a similar improvement over the median to that of the whole sample. For the UK both for production and expenditure the combinations perform similarly to the best model and up to 5% better than the median.

3.4.3 Comments on the Overall Results.

In general terms the results suggest that there are benefits from using forecast combination in this particular case and that most of the gains of doing so are picked up by the equal-weighted scheme. In particular the alternative weighting schemes provide only marginally different results both for the aggregate and the joint combination approaches. Nevertheless the possibility of introducing external knowledge into the combination procedure provides flexibility to the forecaster.

The fact that the aggregate accuracy of the joint combination methods is practically the same, and in some cases marginally better, to that of the equivalent traditional method suggests that the benefits of achieving disaggregate consistency do not come at the cost of the aggregate accuracy.

In fact, given that the combination methods show disaggregate forecasting accuracy similar or better to those of the best performing single models suggests that the constraints that they impose do in fact transmit the benefits of the aggregate methods to the components.

Finding relatively comparable results for three different countries and isolating the direct effect of the financial crisis provides some robustness to the method. This suggests that this method could serve as a valid alternative to the bottom-up only approach.

4 Conclusion

This document develops a framework that incorporates an aggregate and its components into the same forecast combination process. The method performs the combination relying on the merits of the individual forecasts and acknowledges that for any realized outcome an aggregate is exactly the weighted sum of its components. This method makes use of disaggregate components and ensures that the accounting identities that underlie the aggregate are met delivering a completely consistent forecasting scenario.

An important feature of our framework is that it imposes no constraints on the way inn which the forecasts that enter the combination framework are made. Any forecast, based on a model or not, may be used. Given that the method guarantees a consistent scenario, the strengths of the direct aggregate forecast may be transferred to the component's forecasts by effectively constraining the disaggregate set as a whole. This means that some degree of interdependence is forced on the component's forecasts no matter if they are generated independently or not in the first place.

In our empirical application with GDP data from France, Germany and the United Kingdom, we find that our combination framework provides equal aggregate forecasting accuracy to that of equivalent traditional forecast combination methods and disaggregate accuracy similar or better to those of the best performing single models. All this suggests this method is a valid alternative to bottom-up only approaches when a fully consistent scenario is required.

Our method shares many common features with that of Hyndman et al. (2011) but extends it in two aspects. First, it makes it possible to use in contexts where there are multiple forecasts for each variable and multiple distinct possible disaggregations. For example the three different approaches for GDP. Second, it allows to use weights that reflect the relative reliability of the preliminary forecasts themselves. Although in our empirical application the reliability weights determined from the data did not make much of a difference when compared to equal weights, the posibility of establishing the weights coul prove to be useful as a way of introducing external information or judgement into the forecasting process. This is something that Central Banks do regularly as a way of incorporating a broader assessment of relevant conditions that are not explicitly accounted for in their models (Alessi et al., 2014).

A logical step for further research would be to extend our framework to admit multiple levels of disaggregation as is the case with Hyndman et al. (2011). Another immediate area is to explore using it for density forecasting in order to see how it affects the whole distribution. From an applied perspective it would be interesting to enrich the set of models that are included in the combination process. Some obvious candidates would be to add factor models that have boosted the performance of direct aggregate forecasts (Stock and Watson, 1998; Forni et al., 2005) and at the same time incorporate disaggregate methods that include interactions and common features between components within the process (Espasa and Mayo-Burgos, 2013; Esteves, 2013; Stock and Watson, 2015).

Appendix

A Empirical Framework

A.1 Joint Combination with Equal Reliability

Let there be a composite index X that results from the simple sum of $N \geq 2$ strictly positive components x_n. Let there be two forecasts for X for period t. The first, y_t, comes from forecasting X directly, while the second one, Q_t, is the simple sum of the forecasts of its components $q_{n,t}$.

A.1.1 Optimization Weights for One Set of Forecasts

The proportional deviation approach from the reconciliation literature finds the definitive value for X by making the differences between it and the initial estimates proportional. The minimization problem is therefore:

$$\min_{\alpha,\beta} \left[\frac{(1+\alpha)\,y - y}{y}\right]^2 + \left[\frac{(1+\beta)\,Q - Q}{Q}\right]^2 + 2\lambda\left[(1+\alpha)\,y - (1+\beta)\,Q\right] \tag{18}$$

where α and β are the percentage deviations of the definitive value from the initial estimates.[20]

The first order conditions imply that $Q = -\frac{\beta}{\alpha} y$ and $(1+\alpha)\,y = Q + \beta Q$. The aggregate forecast resulting from solving the problem is then:

$$\tilde{y} = (1+\alpha)\,y = (1+\beta)\,Q = \left(\frac{y \cdot Q}{y^2 + Q^2}\right)(y + Q) \tag{19}$$

Using the inequality of arithmetic and geometric means we have that $0 \leq (y - Q)^2 = y^2 + Q^2 - 2yQ$. Then $2yQ \leq y^2 + Q^2$ and therefore:

$$\frac{y \cdot Q}{y^2 + Q^2} \leq \frac{1}{2}$$

meaning that equation (19) is strictly lower than an equal weighted average if both forecasts are distinct.

The additive deviation approach on the other hand finds the definitive value for X by making the differences between it and the initial estimates equal in absolute terms.

[20] Time subscripts are dropped to simplify notation given that all variables are contemporaneous.

The minimization problem can be written as:

$$\min_{\alpha,\beta} \left[(1+\alpha)\,y - y\right]^2 + \left[(1+\beta)\,Q - Q\right]^2 + 2\lambda\left[(1+\alpha)\,y - (1+\beta)\,Q\right] \tag{20}$$

The first order conditions imply that $\beta = -\alpha\frac{y}{Q}$ and $(1+\alpha)y = (1+\beta)Q$. Then replacing β in the latter gives

$$(1+\alpha)y = Q - \alpha y \tag{21}$$

and solving for $(1+\alpha)y$ you get the simple average.

The result does not hold however if the additive approach is used directly on the components forecasts. In that case, the minimization problem is the following:

$$\min_{\alpha,\beta_n} (\alpha y)^2 + \textstyle\sum_{n=1}^{N}(\beta_n q_n)^2 + 2\lambda\left[(1+\alpha)y - \sum_{n=1}^{N}(1+\beta_n)q_n\right] \tag{22}$$

This time the first order conditions imply that $\beta_n = -\alpha\frac{y}{q_n}$ for $n = 1$ to N and $(1+\alpha)y = \sum_{i=1}^{N}(1+\beta_n)q_n$. Solving for $(1+\alpha)\,y$ you get that the aggregate forecast resulting from the combination is:

$$\tilde{y} = \frac{N\cdot y + \sum_{n=1}^{N} q_n}{N+1} = \frac{1}{N+1}\left(N\cdot y + Q\right) \tag{23}$$

that is different to the simple mean given that $N \geq 2$ and both aggregate forecasts are assumed to be distinct.

From comparing both approaches it can be seen that the only difference between the two is that the latter eliminates the downward bias by penalizing deviations based on the relative size of each aggregate forecast. The same idea can be extended to find the appropriate penalty term for the components.

Including an unspecified weight η_n for the disaggregate components in equation (22) results in:

$$\min_{\alpha,\beta_n} (\alpha y)^2 + \sum_{n=1}^{N}(\beta_n \eta_n)^2 + 2\lambda\left[(1+\alpha)y - \sum_{n=1}^{N}(1+\beta_n)q_n\right] \tag{24}$$

This time the first order conditions imply that $\beta_n = -\frac{q_n}{\eta_n^2}\cdot\alpha y$ for $n = 1$ to N and $(1+\alpha)y = \sum_{i=1}^{N}(1+\beta_n)q_n$. Using this gives:

$$(1+\alpha)\,y = \sum_{n=1}^{N}(q_n) - \sum_{n=1}^{N}\left(\frac{q_n^2}{\eta_n^2}\cdot\alpha y\right)$$

Then matching with the intermediate step given by equation (21) results in:

$$Q - \sum_{n=1}^{N}\left(\frac{q_n^2}{\eta_n^2}\cdot\alpha y\right) = Q - \alpha y$$

Then solving for η_n the weight for the components is:

$$\eta_n = \sqrt{q_n \cdot Q}$$

With this, the loss function that produces the equal weighted result for the aggregate is:

$$(\alpha y)^2 + \sum_{n=1}^{N} q_n Q \left(\beta_n\right)^2 \tag{25}$$

A.1.2 Joint Combination for Multiple Distinct Disaggregate Forecasts

The previous framework is not limited to one set of disaggregate forecasts. It can incorporate more forecasts providing they are distinct, that is that they are not based on the same components. To show this let there be a third forecast, G, that is the simple sum of the M forecasts of its components g_m. Again assuming all aggregate forecasts are equally reliable the minimization problem may be written as:

$$\begin{aligned} \min_{\alpha,\beta_n,\gamma_m} \; & (\alpha y)^2 + \textstyle\sum_{n=1}^{N} q_n Q \beta_n^2 + \sum_{m=1}^{M} g_m G \gamma_m^2 \\ & + 2\lambda_1 \left[(1+\alpha)y - \textstyle\sum_{n=1}^{N}(1+\beta_n)q_n\right] + 2\lambda_2 \left[(1+\alpha)y - \textstyle\sum_{m=1}^{M}(1+\gamma_m)g_m\right] \end{aligned} \tag{26}$$

The $N + M + 3$ first order conditions are:

1. $\frac{\partial}{\partial\alpha}: \quad \alpha y + \lambda_1 + \lambda_2 = 0$
2. $\frac{\partial}{\partial\beta_n}: \quad \beta_n Q - \lambda_1 = 0 \qquad$ for $n = 1$ to N
3. $\frac{\partial}{\partial\gamma_m}: \quad \gamma_m G - \lambda_2 = 0 \qquad$ for $m = 1$ to M
4. $\frac{\partial}{\partial\lambda_1}: \quad (1+\alpha)y - \sum_{i=1}^{N}(1+\beta_n)q_n = 0$
5. $\frac{\partial}{\partial\lambda_2}: \quad (1+\alpha)y - \sum_{m=1}^{M}(1+\gamma_m)g_m = 0$

This time the first order conditions imply that $\beta_n = \beta$ for all n and $\gamma_n = \gamma$ for all m. Then 4. and 5. may be rewritten as:

4*. $\quad y + \alpha y = Q + \beta Q$

5*. $\quad y + \alpha y = G + \gamma G$

This means that $\beta Q = y + \alpha y - Q$ and $\gamma G = y + \alpha y - G$ can be replaced directly into 2. and 3. Then plugging these two into 1. gives:

$$\alpha y + (y + \alpha y - Q) + (y + \alpha y - G) = 0$$

and after reordering the expression is:

$$\tilde{y} = \frac{y + Q + G}{3} \tag{27}$$

A.1.3 Bias in Multiple Related Disaggregate Forecast Combination

As shown in section A.1.2 the proposed framework can easily admit more than one set of unrelated forecasts. However, if more than one set of forecasts for the same components are included, a bias similar to that of equation (19) appears. This happens because not only the definitive aggregate forecasts have to coincide, but also those of the components.

Extending the framework in equation (25) to a setting with D sets of disaggregate forecasts for the N components the minimization problem may be written as:

$$\begin{aligned}
&\min_{\alpha,\beta_n} (\alpha y)^2 + \textstyle\sum_{d=1}^{D} \sum_{n=1}^{N} q_{d,n} Q_d \beta_{d,n}^2 \\
&\quad + 2 \textstyle\sum_{d=1}^{D} \left(\lambda_d \left[(1+\alpha) y - \sum_{n=1}^{N} (1+\beta_{d,n}) q_{d,n} \right] \right) \\
&\quad + 2 \textstyle\sum_{d=2}^{D} \sum_{n=1}^{N} \left(\delta_{d,n} \left[(1+\beta_{1,n}) q_{1,n} - (1+\beta_{d,n}) q_{d,n} \right] \right)
\end{aligned} \tag{28}$$

Simplifying the problem to the particular case with one aggregate, two disaggregate forecasts and $N = 2$ the first order conditions become:

1. $\frac{\partial}{\partial \alpha}: \quad \alpha y + \lambda_1 + \lambda_2 = 0$
2. $\frac{\partial}{\partial \beta_{1,n}}: \quad \beta_{1,n} Q_1 - \lambda_1 + \delta_n = 0$ for $n = 1, 2$
3. $\frac{\partial}{\partial \beta_{2,n}}: \quad \beta_{2,n} Q_2 - \lambda_2 - \delta_n = 0$ for $n = 1, 2$
4. $\frac{\partial}{\partial \lambda_d}: \quad (1+\alpha) y - (1+\beta_{d,1}) q_{d,1} - (1+\beta_{d,2}) q_{d,2} = 0$ for $d = 1, 2$
5. $\frac{\partial}{\partial \delta_n}: \quad (1+\beta_{1,n}) q_{1,n} - (1+\beta_{2,n}) q_{2,n} = 0$ for $n = 1, 2$

After some algebra using conditions 1, 2, 3 and 5 we have $(1+\beta_{1,n}) = q_{2,n}(Q_1 q_{2,n} + Q_2 q_{1,n})^{-1}(Q_1 + Q_2 - \alpha y)$ for $n = 1, 2$. Using this in the corresponding condition in 4. we get:

$$\begin{aligned}
\tilde{y} &= \Phi (Q_1 + Q_2 - \alpha y) \\
&= \tfrac{\Phi}{1+\Phi} (y + Q_1 + Q_2)
\end{aligned} \tag{29}$$

where

$$\Phi = \frac{Q_1^2 q_{2,1} q_{2,2} + Q_2^2 q_{1,1} q_{1,2}}{Q_1^2 q_{2,1} q_{2,2} + Q_1 Q_2 (q_{1,2} q_{2,1} + q_{1,1} q_{2,2}) + Q_2^2 q_{1,1} q_{1,2}}$$

For equation (29) to be the simple average it is necessary for $\frac{1+\Phi}{\Phi}$ to be equal to three. This is equivalent to saying that $\Phi^{-1} - 1$, that is given by:

$$\Phi^{-1} - 1 = \frac{Q_1 Q_2 (q_{1,2} q_{2,1} + q_{1,1} q_{2,2})}{Q_1^2 q_{2,1} q_{2,2} + Q_2^2 q_{1,1} q_{1,2}}$$

has to be equal to one.

To explore under what circumstances this is in fact true, we express the second set of preliminary estimates as deviations from the first set, that is $q_{2,1} = \kappa_1 q_{1,1}$ and $q_{2,2} = \kappa_2 q_{1,2}$ where κ_1 and κ_2 can take any value. Assuming that $\Phi^{-1} - 1$ is in fact equal to one we have:

$$\frac{Q_1 (\kappa_1 q_{1,1} + \kappa_2 q_{1,2})(\kappa_1 q_{1,1} q_{1,2} + q_{1,1} \kappa_2 q_{1,2})}{Q_1^2 \kappa_1 \kappa_2 q_{1,1} q_{1,2} + (\kappa_1 q_{1,1} + \kappa_2 q_{1,2})^2 q_{1,1} q_{1,2}} = 1$$

Then:

$$(q_{1,1} + q_{1,2})(\kappa_1 q_{1,1} + \kappa_2 q_{1,2})(\kappa_1 + \kappa_2) = Q_1^2 \kappa_1 \kappa_2 + (\kappa_1 q_{1,1} + \kappa_2 q_{1,2})^2$$

$$(\kappa_1 q_{1,1} + \kappa_2 q_{1,2})(\kappa_1 q_{1,2} + \kappa_2 q_{1,1}) = Q_1^2 \kappa_1 \kappa_2$$

$$\kappa_1^2 q_{1,1} q_{1,2} + \kappa_2^2 q_{1,1} q_{1,2} = 2 \kappa_1 \kappa_2 q_{1,1} q_{1,2}$$

$$\kappa_1^2 - 2\kappa_1 \kappa_2 + \kappa_2^2 = 0$$

that results in $(\kappa_1 - \kappa_2)^2 = 0$.

This condition only holds when $\kappa_1 = \kappa_2 = \kappa$ meaning that the outcome of equation (28) is a simple average only when the two sets of preliminary estimates are exactly the same or the second one is simply the first multiplied by a constant.

A.2 Joint Combination using Reliability Weights

For a composite index X that results from the weighted sum of $N \geq 2$ strictly positive components x_n. Let there be $A+D$ forecasts for X in period $t+h$. The first forecasts, y_i for $i = 1$ to A, come from forecasting the aggregate X directly, while the remaining, Q_d for $d = 1$ to D, are each a weighted sum of the respective set of components forecasts $q_{d,n}$.[21] The aggregation weights are assumed to be positive.

A.2.1 Combination for One Set of Forecasts

Let there be a single aggregate forecast y and a single set of disaggregate forecasts q_n for $n = 1$ to N. The minimization problem involving the aggregate reliability weight φ, the disaggregate reliability weights ϕ_n and the aggregation weights w_n, is:

$$\min_{\alpha,\beta} \varphi\left(\alpha y\right)^2 + Q\sum_{n=1}^{N} \phi_n w_n q_n \left(\beta_n\right)^2 + 2\lambda\left[(1+\alpha)y - \sum_{n=1}^{N} w_n(1+\beta_n)q_n\right] \tag{30}$$

The $N+2$ first order conditions imply that $\beta_n = -\alpha y \frac{\varphi}{\phi_n}\frac{1}{Q}$ and $(1+\alpha)y = \sum_{n=1}^{N} w_n(1+\beta_n)q_n$. Combining these mean that:

$$y + \alpha y = Q + \textstyle\sum_{n=1}^{N} \beta_n w_n q_n$$

$$y + \alpha y = Q - \tfrac{\alpha y}{Q}\textstyle\sum_{n=1}^{N} \tfrac{\varphi}{\phi_n} w_n q_n$$

$$Qy + \alpha y\left(Q + \textstyle\sum_{n=1}^{N} \tfrac{\varphi}{\phi_n} w_n q_n\right) = Q^2$$

$$(1+\alpha)\, y\left(Q + \textstyle\sum_{n=1}^{N} \tfrac{\varphi}{\phi_n} w_n q_n\right) = Q^2 + y\textstyle\sum_{n=1}^{N} \tfrac{\varphi}{\phi_n} w_n q_n$$

Then finally the outcome for the aggregate forecast is a weighted average:

$$\tilde{y} = \frac{Q^2 + y\sum_{n=1}^{N} \frac{\varphi}{\phi_n} w_n q_n}{Q + \sum_{n=1}^{N} \frac{\varphi}{\phi_n} w_n q_n} \tag{31}$$

The components are then immediately obtainable from α and the first order conditions. Then:

$$\tilde{q}_n = \left(1 + \frac{\varphi}{\phi_n}\frac{y - Q}{Q + \sum_{n=1}^{N} \frac{\varphi}{\phi_n} w_n q_n}\right) q_n \tag{32}$$

[21] For simplicity the framework is derived for one set of components but as shown in equation (27) additional sets of components may be added easily.

If y and q_n for $n = 1$ to N are the result of a prior combination, that is $y = \frac{1}{\Gamma}\sum_{i=1}^{A}\gamma_i y_i$ and $q_n = \frac{1}{\Delta_n}\sum_{d=1}^{D}\delta_{d,n}q_{d,n}$ with γ_i and $\delta_{d,n}$ being the prior reliability weights, $\Gamma = \sum_{i=1}^{A}\gamma_i$ and $\Delta_n = \sum_{d=1}^{D}\delta_{d,n}$, the equivalence with the simple average of the initial forecasts can be shown by replacing them into the solution:

$$\tilde{y} = \frac{\left(\sum_{n=1}^{N} w_n \sum_{d=1}^{D}\frac{\delta_{d,n}}{\Delta_n}q_{d,n}\right)^2 + \left(\frac{1}{\Gamma}\sum_{i=1}^{A}\gamma_i y_i\right)\left(\sum_{n=1}^{N}\frac{\varphi}{\phi_n}w_n\sum_{d=1}^{D}\frac{\delta_{d,n}}{\Delta_n}q_{d,n}\right)}{\left(\sum_{n=1}^{N} w_n \sum_{d=1}^{D}\frac{\delta_{d,n}}{\Delta_n}q_{d,n}\right) + \left(\sum_{n=1}^{N}\frac{\varphi}{\phi_n}w_n\sum_{d=1}^{D}\frac{\delta_{d,n}}{\Delta_n}q_{d,n}\right)} \tag{33}$$

Then incorporating the equal reliability of forecasts by setting $\gamma_i = \delta_{d,n,t} = 1$ and reflecting the number of forecasts that are involved in the first stage with $\varphi = A$ and $\phi_n = \phi = D$. The solutions then simplifies as follows:

$$\begin{aligned}
\tilde{y} &= \frac{\left(\frac{1}{D}\sum_{d=1}^{D}\sum_{n=1}^{N} w_n q_{d,n}\right)^2 + \left(\frac{1}{A}\sum_{i=1}^{A} y_i\right)\left(\frac{A}{D}\cdot\frac{1}{D}\sum_{d=1}^{D}\sum_{n=1}^{N} w_n q_{d,n}\right)}{\left(\frac{1}{D}\sum_{d=1}^{D}\sum_{n=1}^{N} w_n q_{d,n}\right) + \left(\frac{A}{D}\cdot\frac{1}{D}\sum_{d=1}^{D}\sum_{n=1}^{N} w_n q_{d,n}\right)} \\
&= \frac{\left(\frac{1}{D}\sum_{d=1}^{D}\sum_{n=1}^{N} w_n q_{d,n}\right) + \left(\frac{A}{D}\sum_{i=1}^{I} y_i\right)}{1+\frac{A}{D}} \\
&= \frac{1}{A+D}\left(\sum_{i=1}^{A} y_i + \sum_{d=1}^{D} Q_d\right)
\end{aligned}$$

that is the simple average of all the aggregate forecasts.

A.2.2 Combination for Multiple Distinct Disaggregate Forecasts

In a way similar to that of section A.1.2 the developed framework is not limited to one set of disaggregate forecasts. It can incorporate more forecasts providing the are distinct, that is that they are not based on the same components. Building on the the set-up in section A.2.1, additional to the direct aggregate forecast y, let there by K distinct aggregate forecasts each based on the corresponding N_k component's forecasts.

The minimization problem involving the aggregate reliability weight φ, the disaggregate reliability weights $\phi_{k,n}$ and the aggregation weights $w_{k,n}$, is:

$$\min_{\alpha,\beta} \varphi\,(\alpha y)^2 + \sum_{k=1}^{K}\left[Q_k \sum_{n=1}^{N_k}\phi_{k,n}w_{k,n}q_{k,n}\,(\beta_{k,n})^2 + 2\lambda_k\left((1+\alpha)y - \sum_{n=1}^{N_k} w_{k,n}(1+\beta_{k,n})q_{k,n}\right)\right]$$

The first order conditions are:

1. $\frac{\partial}{\partial\alpha}:\quad \varphi\alpha y + \sum_{k=1}^{K}\lambda_k = 0$

2. $\frac{\partial}{\partial \beta_{k,n}}$: $Q_k \phi_{k,n} \beta_{k,n} - \lambda_k = 0$ for $n = 1$ to N_k and $k = 1$ to K

3. $\frac{\partial}{\partial \lambda_k}$: $(1+\alpha)y - \sum_{n=1}^{N_k} w_{k,n}(1+\beta_{k,n})q_{k,n} = 0$

From 2. and for any k we have that $\phi_{k,n}\beta_{k,n} = \frac{\lambda_k}{Q_k}$ and plugging in the corresponding restriction in 3. we get:

$$(1+\alpha)y = \sum_{n=1}^{N_k} w_{k,n}q_{k,n} + \sum_{n=1}^{N_k} w_{k,n}\beta_{k,n}q_{k,n}$$

$$y + \alpha y = Q_k + \frac{\lambda_k}{Q_k} \sum_{n=1}^{N_k} \left(\frac{1}{\phi_{k,n}} w_{k,n} q_{k,n} \right)$$

$$\lambda_k = \left[\sum_{n=1}^{N_k} \frac{1}{\phi_{k,n}} w_{k,n} q_{k,n} \right]^{-1} Q_k \left(y + \alpha y - Q_k \right)$$

Then using 1. and dividing by φ we get:

$$\begin{aligned} \alpha y &= \sum_{k=1}^{K} \left(\left[\sum_{n=1}^{N_k} \frac{\varphi}{\phi_{k,n}} w_{k,n} q_{k,n} \right]^{-1} Q_k \left(Q_k - y - \alpha y \right) \right) \\ &= \sum_{k=1}^{K} \frac{Q_k}{\chi_k} \left(Q_k - y - \alpha y \right) \end{aligned}$$

where $\chi_k = \sum_{n=1}^{N_k} \frac{\varphi}{\phi_{k,n}} w_{k,n} q_{k,n}$.

The previous equations can be manipulated as follows:

$$\alpha y = \sum_{k=1}^{K} \frac{Q_k}{\chi_k} Q_k - \sum_{k=1}^{K} \frac{Q_k}{\chi_k} y - \sum_{k=1}^{K} \frac{Q_k}{\chi_k} \alpha y$$

$$\left(1 + \sum_{k=1}^{K} \frac{Q_k}{\chi_k} \right) \alpha y = \sum_{k=1}^{K} \frac{Q_k}{\chi_k} Q_k - \left(1 + \sum_{k=1}^{K} \frac{Q_k}{\chi_k} \right) y + y$$

Then the definitive aggregate forecast is seen to be a weighted average given by:

$$\tilde{y} = \frac{y + \sum_{k=1}^{K} \left(Q_k \cdot \frac{Q_k}{\chi_k} \right)}{1 + \sum_{k=1}^{K} \frac{Q_k}{\chi_k}} \tag{34}$$

The definitive component forecasts are obtained combining 2. and λ_k. Then:

$$Q_k \phi_{k,n} \beta_{k,n} - \varphi \frac{Q_k}{\chi_k} \left(y + \alpha y - Q_k \right) = 0$$

$$\beta_{k,n} = \frac{\varphi}{\phi_{k,n}} \frac{Q_k}{\chi_k} \left(y + \alpha y - Q_k \right) \frac{1}{Q_k}$$

with the final result being:

$$\tilde{q}_{k,n} = \left(1 + \frac{\varphi}{\phi_{k,n}} \cdot \frac{\tilde{y} - Q_k}{\chi_k}\right) q_{k,n} \tag{35}$$

A.2.3 Optimal Weights for Multiple Distinct Disaggregate Forecasts

Hyndman et al. (2011) propose a method for obtaining consistent forecasts for a whole hierarchy of time series using a regression approach. The data is described by

$$\mathbf{Y}_t = \mathbf{S}\mathbf{Y}_{K,t} \tag{36}$$

where $\mathbf{Y}_t$ is a vector containing the values for all the series in the hierarchy at time t, $\mathbf{S}$ is the aggregation matrix that defines the structure of the hierarchy and $\mathbf{Y}_{K,t}$ is a vector containing the values at time t for the series at the lowest level of the hierarchy (maximum disaggregation).

For a hierarchy composed of four components, two intermediate aggregations and the total, for example, the vector for lowest level would be $\mathbf{Y}_{2,t} = \begin{bmatrix} y_{2,1,t} & y_{2,2,t} & y_{2,3,t} & y_{2,4,t} \end{bmatrix}'$, the vector for all observations would be $\mathbf{Y}_t = \begin{bmatrix} y_{0,t} & y_{1,1,t} & y_{1,2,t} & \mathbf{Y}'_{2,t} \end{bmatrix}'$ and the aggregation matrix would be:

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}'$$

Hyndman et al. (2011) propose using this same structure to find consistent definitive forecasts from a set of independent forecasts for all series. They set up the following problem for the forecasts at time h:

$$\widetilde{\mathbf{Y}}_h = \mathbf{S}\mathbf{P}\hat{\mathbf{Y}}_h \tag{37}$$

where $\hat{\mathbf{Y}}_h$ are the preliminary forecasts for all series, $\widetilde{\mathbf{Y}}_h$ are the consistent definitive forecasts for all series and $\mathbf{P}$ is a balancing matrix. They use the regression approach to find $\mathbf{P}$ and in particular assume that the forecast errors satisfy the same aggregation constraint as the data (36). Under these assumptions they find that the optimal balancing matrix is $\mathbf{P} = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ and that therefore

$$\widetilde{\mathbf{Y}}_h = \mathbf{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}'\hat{\mathbf{Y}}_h \tag{38}$$

For a one level hierarchical structure with N components with $\hat{\mathbf{Y}}_t = \begin{bmatrix} y_t & q_1 & \cdots & q_N \end{bmatrix}'$

we have that

$$\mathbf{S}\left(\mathbf{S}'\mathbf{S}\right)^{-1}\mathbf{S}' = \frac{1}{(N+1)}\begin{bmatrix} N & 1 & 1 & \cdots & 1 \\ 1 & N & -1 & \cdots & -1 \\ 1 & -1 & N & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ 1 & -1 & \cdots & -1 & N \end{bmatrix}$$

The outcome of using equation (38) is then

$$\widetilde{y} = \frac{1}{N+1}\left(N \cdot y + \sum_{n=1}^{N} q_n\right)$$

and

$$\widetilde{q_n} = q_n + \frac{1}{N+1}\left(y - \sum_{n=1}^{N} q_n\right)$$

As it can be seen from comparing this solution to the additive maximization problem in equation (23), the results are the same. It is only a matter of setting the reliability weights in equation (12) to $\varphi_t = 1$ and $\phi_{n,t} = q_{n,t}/Q_t$ to arrive at the standard additive maximization problem and therefore reproducing the optimal combination method.

The optimal combination method by Hyndman et al. (2011), however, does not contemplate multiple distinct sets of disaggregate forecast. However, by using the similarity with the additive minimization problem we can extend the idea to do so. Using the same definitions as in equation (12) and setting the respective weights, $\varphi = 1$, $\phi_n = \frac{q_n}{\sum_{n=1}^{N} q_n}$ and $\psi_m = \frac{g_m}{\sum_{m=1}^{M} g_m}$, the minimization problem is the following:

$$\begin{aligned} \min_{\alpha,\beta_n,\gamma_m} \; & (\alpha y)^2 + \textstyle\sum_{n=1}^{N} (\beta_n q_n)^2 + \sum_{m=1}^{M} (\gamma_m g_m)^2 \\ & + 2\lambda_1 \left[(1+\alpha)y - \textstyle\sum_{n=1}^{N}(1+\beta_n)q_n\right] \\ & + 2\lambda_2 \left[(1+\alpha)y - \textstyle\sum_{m=1}^{M}(1+\gamma_m)g_m\right] \end{aligned} \tag{39}$$

The $N + M + 3$ first order conditions are:

1. $\frac{\partial}{\partial \alpha}: \quad \alpha y + \lambda_1 + \lambda_2 = 0$
2. $\frac{\partial}{\partial \beta_n}: \quad \beta_n q_n - \lambda_1 = 0$ for $n = 1$ to N
3. $\frac{\partial}{\partial \gamma_m}: \quad \gamma_m g_m - \lambda_2 = 0$ for $m = 1$ to M
4. $\frac{\partial}{\partial \lambda_1}: \quad (1+\alpha)y - \sum_{i=1}^{N}(1+\beta_n)q_n = 0$
5. $\frac{\partial}{\partial \lambda_2}: \quad (1+\alpha)y - \sum_{m=1}^{M}(1+\gamma_m)g_m = 0$

By substituting 2. and 3. into 4. and 5. respectively we have that:

4*. $(1+\alpha)y = Q + N\lambda_1$

5*. $(1+\alpha)y = G + M\lambda_2$

where $Q = \sum_{n=1}^{N} q_n$ and $G = \sum_{m=1}^{M} g_m$. Then solving for λ_1 and λ_2 and replacing in 1. gives:

$$\alpha y + \frac{1}{N}\left(y + \alpha y - Q\right) + \frac{1}{M}\left(y + \alpha y - G\right) = 0$$

and after reordering the expression for the aggregate is:

$$\tilde{y}^a = \frac{1}{NM + M + N}\left(NM \cdot y + M \cdot Q + N \cdot G\right) \tag{40}$$

Regarding the disaggregates we have from 2. and 4*. that:

$$\beta_n q_n = \tfrac{(1+\alpha)y - Q}{N} \quad \Rightarrow \quad \tilde{q}_n^a = q_n + \tfrac{\tilde{y}^a - Q}{N}$$

and from 3. and 5*. that:

$$\gamma_m g_m = \tfrac{(1+\alpha)y - G}{M} \quad \Rightarrow \quad \tilde{g}_m^a = g_m + \tfrac{\tilde{y}^a - G}{M}$$

B Empirical Application

B.1 Single Model Aggregate Relative Forecasting Errors

B.1.1 Whole Sample

Table 8: Single Model Aggregate Relative Forecasting Errors by Approach

	Direct				Production				Expenditure			
Horizon	1	2	3	4	1	2	3	4	1	2	3	4
France												
RW	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-
AR	0.86	0.90	0.95	0.99	0.87	0.90	0.94	0.97	0.88	0.91	0.95	0.98
ARIMA	0.86	0.90	1.02	1.10	0.90	0.94	1.02	1.08	0.90	0.97*	1.03	1.14*
SVDIF	0.85	0.86	0.89	0.91	0.92**	0.95**	1.00**	1.04**	0.91*	0.90	0.92	0.93
SVDDIF	0.84	0.83	0.85	0.86	0.94**	0.97**	1.03**	1.07**	0.93**	0.94**	0.97	0.98
LVLEV	0.93*	0.96**	0.99**	1.02**	0.93*	0.96**	0.99**	1.02**	0.93*	0.96**	0.99**	1.01**
LVDIF	0.93**	0.95**	1.00**	1.03**	0.92**	0.94**	0.99**	1.02**	0.92**	0.95**	0.99**	1.02**
LVDDIF	0.95**	0.98**	1.03**	1.06**	0.94**	0.97**	1.02**	1.06**	0.94**	0.97**	1.02**	1.06**
MIN	0.84	0.83	0.85	0.86	0.87	0.90	0.94	0.97	0.88	0.90	0.92	0.93
MAX	1.00	1.00	1.03	1.10	1.00	1.00	1.03	1.08	1.00	1.00	1.03	1.14
MEDIAN	0.90	0.93	0.99	1.01	0.93	0.96	1.00	1.03	0.93	0.95	0.99	1.00
Germany												
RW	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-
AR	1.02	1.09	1.13	1.16	0.93	0.99	1.03	1.04	1.04	1.09	1.11	1.12
ARIMA	1.00	1.07	1.13	1.18	1.00	1.12*	1.13	1.13	1.06	1.28	1.37	1.40
SVDIF	0.94	0.96	0.97	0.99	1.01	1.04	1.06	1.09	0.97	1.03	1.07	1.10
SVDDIF	0.95	0.96	0.97	0.99	1.03	1.04	1.06	1.08	0.98	1.02	1.06	1.09
LVLEV	1.00	1.05*	1.09*	1.14*	1.00	1.04*	1.09*	1.14*	1.01*	1.05*	1.09*	1.14*
LVDIF	1.00	1.06	1.09	1.12	0.98	1.03	1.07	1.09	0.99	1.03	1.06	1.09
LVDDIF	1.02	1.07	1.10	1.13	1.00	1.04	1.07	1.10	1.00	1.04	1.06	1.09
MIN	0.94	0.96	0.97	0.99	0.93	0.99	1.01	1.01	0.97	1.01	1.02	1.02
MAX	1.02	1.09	1.13	1.18	1.03	1.12	1.13	1.14	1.06	1.28	1.37	1.40
MEDIAN	1.00	1.05	1.09	1.12	1.00	1.04	1.06	1.09	1.00	1.04	1.07	1.09
UK												
RW	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-
AR	0.75	0.84	0.94	0.99	0.89	0.93	0.96	0.98	0.97	0.96	0.97	0.98
ARIMA	0.77	0.89	1.02	1.11	0.92*	0.95	0.97	0.99	1.13**	1.11**	1.11**	1.10**
SVDIF	0.72	0.80	0.88	0.92	0.76	0.86	0.94	0.97	0.78	0.86	0.90	0.95
SVDDIF	0.71	0.78	0.87	0.91	0.76	0.85	0.93	0.96	0.77	0.86	0.90	0.94
LVLEV	0.90	0.94	0.98	1.01	0.90	0.94	0.98	1.01	0.90	0.94	0.98	1.01
LVDIF	0.71	0.82	0.89	0.94	0.72	0.82	0.90	0.94	0.71	0.82	0.89	0.93
LVDDIF	0.71	0.81	0.88	0.93	0.71	0.81	0.89	0.93	0.71	0.81	0.88	0.92
MIN	0.71	0.78	0.87	0.91	0.71	0.81	0.89	0.93	0.71	0.81	0.88	0.92
MAX	1.00	1.00	1.02	1.11	1.01	1.01	1.01	1.02	1.13	1.11	1.11	1.10
MEDIAN	0.73	0.83	0.92	0.96	0.83	0.90	0.95	0.97	0.84	0.90	0.94	0.96

Note: Mean square forecasting error of each model relative to that of the direct approach using the random walk model for each horizon. The models are a random walk with drift (RW), a first-differences autoregressive model of order one (AR), an ARIMA chosen according to the Bayesian Information Criterion (ARIMA), two small VARs including CPI and the GDP variables from each approach in first differences (SVDIF) and where CPI is differenced twice (SVDDIF) and three large VARs including CPI and all GDP variables in levels (LVLEV), in first differences (LVDIF) and in first differences with CPI differenced twice (LVDDIF). * and ** denote that the respective forecast is statistically worse than the best model for that country according to the Modified Diebold-Mariano statistic at a 10 and 5% significance level. Calculated for one to four steps ahead forecasts over the 2001-2014 period.

B.1.2 Restricted Sample

Table 9: Single Model Forecasting Errors excluding 2008-2009

	Direct				Production				Expenditure			
Horizon	1	2	3	4	1	2	3	4	1	2	3	4
France												
RW	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-
AR	1.06	1.03	1.04	1.06**	1.00	0.97	1.01	1.06	1.05	1.01	1.01	1.03
ARIMA	1.05	1.05	1.17	1.22	1.00	0.95	1.02	1.06	1.23	1.31	1.36	1.52*
SVDIF	1.05	1.05	1.09	1.12	1.13	1.23*	1.37**	1.50**	1.10	1.09	1.12*	1.14**
SVDDIF	1.06*	1.06	1.10	1.13	1.17*	1.27**	1.42**	1.57**	1.14*	1.13	1.15**	1.19**
LVLEV	1.14**	1.26**	1.35**	1.42**	1.14**	1.26*	1.35**	1.43**	1.14**	1.26**	1.35**	1.42**
LVDIF	1.12	1.21*	1.35**	1.47**	1.12	1.20	1.34**	1.47**	1.13	1.23*	1.37**	1.50**
LVDDIF	1.16*	1.25*	1.42**	1.56**	1.16*	1.24*	1.41**	1.55**	1.17*	1.27*	1.44**	1.58**
MIN	1.00	1.00	1.00	1.00	1.00	0.95	1.01	1.01	1.01	1.01	1.01	1.03
MAX	1.16	1.26	1.42	1.56	1.17	1.27	1.42	1.57	1.23	1.31	1.44	1.58
MEDIAN	1.06	1.06	1.13	1.17	1.13	1.22	1.35	1.45	1.13	1.18	1.25	1.31
Germany												
RW	1.00*	1.00*	1.00	1.00	-	-	-	-	-	-	-	-
AR	0.96	0.94*	0.93	0.94**	0.87	0.90	0.93	0.92	0.95	0.94	0.91	0.92
ARIMA	1.01**	0.98**	0.98	0.98	1.10**	1.28*	1.24	1.17	1.02	1.17*	1.19	1.16
SVDIF	0.98	1.00	1.02	1.04	1.03	1.05	1.06	1.08	0.94	0.96	0.98	0.96
SVDDIF	0.99*	0.98*	0.98	1.00	1.11*	1.09*	1.08	1.10	0.99*	1.00	1.00	0.99
LVLEV	1.03*	1.08**	1.13*	1.17**	1.03*	1.09**	1.14*	1.18**	1.04*	1.09**	1.14*	1.18**
LVDIF	0.96	0.99	1.01	1.03	0.96	1.00	1.01	1.03	0.97	1.00	1.01	1.02
LVDDIF	1.00	1.01	1.02	1.03	1.00	1.02	1.03	1.04	1.01	1.02	1.03	1.03
MIN	0.96	0.94	0.93	0.94	0.87	0.90	0.93	0.92	0.94	0.94	0.91	0.92
MAX	1.03	1.08	1.13	1.17	1.11	1.28	1.24	1.18	1.04	1.17	1.19	1.18
MEDIAN	1.00	0.99	1.01	1.01	1.02	1.03	1.04	1.06	0.99	1.00	1.01	1.00
UK												
RW	1.00	1.00	1.00	1.00	-	-	-	-	-	-	-	-
AR	1.11	1.05	1.10	1.06	1.05	0.97	0.98	0.98	1.19	1.12*	1.04	1.04
ARIMA	1.18*	1.11	1.17	1.14	1.18*	1.14*	1.15**	1.21**	1.89**	2.12**	2.12**	2.20**
SVDIF	1.12	1.20**	1.35**	1.43**	1.08	1.19*	1.38**	1.47**	1.15	1.27*	1.25*	1.34*
SVDDIF	1.14	1.18**	1.31**	1.37**	1.12	1.21*	1.39**	1.47**	1.15	1.26**	1.20	1.27
LVLEV	1.16	1.29*	1.42*	1.55*	1.16	1.29*	1.41*	1.53*	1.15	1.28*	1.39*	1.52*
LVDIF	1.08	1.22*	1.36**	1.46**	1.09	1.22*	1.35**	1.45**	1.05	1.21*	1.34**	1.44**
LVDDIF	1.10	1.22*	1.35**	1.44**	1.10	1.22*	1.34**	1.43**	1.08	1.21*	1.34**	1.43**
MIN	1.00	1.00	1.00	1.00	1.05	0.97	0.98	0.98	1.05	1.12	1.04	1.04
MAX	1.18	1.29	1.42	1.55	1.18	1.29	1.41	1.53	1.89	2.12	2.12	2.20
MEDIAN	1.12	1.19	1.33	1.40	1.10	1.21	1.35	1.45	1.15	1.26	1.34	1.43

Note: Mean square forecasting error of each model relative to that of the direct approach using the random walk model for each horizon. The models are a random walk with drift (RW), a first-differences autoregressive model of order one (AR), an ARIMA chosen according to the Bayesian Information Criterion (ARIMA), two small VARs including CPI and the GDP variables from each approach in first differences (SVDIF) and where CPI is differenced twice (SVDDIF) and three large VARs including CPI and all GDP variables in levels (LVLEV), in first differences (LVDIF) and in first differences with CPI differenced twice (LVDDIF). * and ** denote that the respective forecast is statistically worse than the best model for that country according to the Modified Diebold-Mariano statistic at a 10 and 5% significance level. Calculated for one to four steps ahead forecasts over the 2001-2014 period excluding years 2008 and 2009.

References

Aiolfi, M. and C. A. Favero (2005). Model uncertainty, thick modelling and the predictability of stock returns. *Journal of Forecasting 24*(4), 233–254.

Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central bank macroeconomic forecasting during the global financial crisis: the european central bank and federal reserve bank of new york experiences. *Journal of Business & Economic Statistics 32*(4), 483–500.

Banbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*, 451–468.

Benalal, N., J. L. Diaz del Hoyo, B. Landau, M. Roma, and F. Skudelny (2004). To aggregate or not to aggregate? euro area inflation forecasting. Working Paper 374, European Central Bank.

Brüggemann, R. and H. Lütkepohl (2013). Forecasting contemporaneous aggregates with stochastic aggregation weights. *International Journal of Forecasting 29*(1), 60–68.

Burriel, P. (2012). A real-time disaggregated forecasting model for the euro area gdp. *Economic Bulletin*, 93–103.

Chauvet, M. and S. Potter (2013). Forecasting output. *Handbook of Economic Forecasting 2*(Part A), 141–194.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting 5*(4), 559–583.

Conflitti, C., C. De Mol, and D. Giannone (2015). Optimal combination of survey forecasts. *International Journal of Forecasting 31*(4), 1096–1103.

Dalgaard, E. and C. Gysting (2004). An algorithm for balancing commodity-flow systems. *Economic Systems Research 16*(2), 169–190.

Diebold, F. X. and J. A. Lopez (1996). Forecast evaluation and combination. in Maddala and Rao, eds., Handbook of Statistics (Elsevier, Amsterdam).

Drechsel, K. and R. Scheufele (2013). Bottom-up or direct? forecasting german gdp in a data-rich environment. IWH Discussion Papers 7, Halle Institute for Economic Research.

ECB (2015). The ECB survey of professional forecasters 2nd Quarter of 2015. European Central Bank. April 2015.

Eklund, J. and S. Karlsson (2007). Forecast combination and model averaging using predictive measures. *Econometric Reviews 26*(2-4), 329–363.

Espasa, A. and I. Mayo-Burgos (2013). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting 29*(4), 718–732.

Espasa, A., E. Senra, and R. Albacete (2002). Forecasting inflation in the european monetary union: A disaggregated approach by countries and by sectors. *The European Journal of Finance 8*(4), 402–421.

Esteves, P. S. (2013). Direct vs bottom–up approach when forecasting gdp: Reconciling literature results with institutional practice. *Economic Modelling 33*, 416–420.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model. *Journal of the American Statistical Association 100*(471).

Garcia, J. A. (2003). An introduction to the ecb's survey of professional forecasters. *ECB Occasional Paper* (8).

Giacomini, R. and C. W. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics 118*(1), 7–26.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014). Short-term inflation projections: A bayesian vector autoregressive approach. *International Journal of Forecasting 30*(3), 635–644.

Gomez, V. and A. Maravall (1996). Programs TRAMO and SEATS. Instructions for the User. Working paper, Banco de Espana. 9628, Research Department, Bank of Spain.

Granger, C. W. (1987). Implications of aggregation with common factors. *Econometric Theory 3*(02), 208–222.

Granger, C. W. and R. Ramanathan (1984). Improved methods of combining forecasts. *Journal of Forecasting 3*(2), 197–204.

Hahn, E. and F. Skudelny (2008). Early estimates of euro area real gdp growth: a bottom up approach from the production side. Working Paper Series 0975, European Central Bank.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics 146*(2), 342–350.

Hendry, D. F. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business & Economic Statistics 29*(2).

Hsiao, C. and S. K. Wan (2014). Is there an optimal forecast combination? *Journal of Econometrics 178*, 294–309.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by hicp component improve forecast accuracy? *International Journal of Forecasting 21*(1), 119–136.

Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics & Data Analysis 55*(9), 2579–2589.

Kapetanios, G., V. Labhard, and S. Price (2008). Forecasting using bayesian and information-theoretic model averaging: an application to uk inflation. *Journal of Business & Economic Statistics 26*(1), 33–41.

Lequiller, F. and D. Blades (2014). *Understanding National Accounts: Second Edition*. OECD Publishing.

Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*, Volume 284. Springer Science & Business Media.

Lütkepohl, H. (2011). Forecasting nonlinear aggregates and aggregates with time-varying weights. *Jahrbücher für Nationalökonomie und Statistik*, 107–133.

Marcellino, M. (2008). A linear benchmark for forecasting gdp growth and inflation? *Journal of Forecasting 27*(4), 305–340.

Marcellino, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the euro area: Country specific versus area-wide information. *European Economic Review 47*(1), 1–18.

Newbold, P. and D. I. Harvey (2002). Forecast combination and encompassing. M.P. Clements and D.F. Hendry, eds., A Companion to Economic Forecasting (Blackwell, Oxford), 268–283.

OECD (2009). The situation of quarterly national accounts data transmission to the OECD. Working Party on National Accounts WPNA(2009)1. OECD.

Perevalov, N. and P. Maier (2010). On the advantages of disaggregated data: insights from forecasting the us economy in a data-rich environment. Working Papers 10-10, Bank of Canada.

Ravazzolo, F. and S. P. Vahey (2014). Forecast densities for economic aggregates from disaggregate ensembles. *Studies in Nonlinear Dynamics & Econometrics 18*(4), 367–381.

Rodrigues, J. F. (2014). A bayesian approach to the balancing of statistical economic data. *Entropy 16*(3), 1243–1271.

Smith, J. and K. F. Wallis (2009). A simple explanation of the forecast combination puzzle*. *Oxford Bulletin of Economics and Statistics 71*(3), 331–355.

Stock, J. and M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. in R.F. Engle and H. White, eds., Festschrift in Honour of Clive Granger (Cambridge University Press, Cambridge) 1-44.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. Working Paper 6702, NBER.

Stock, J. H. and M. W. Watson (2015). Core inflation and trend inflation. Technical report, National Bureau of Economic Research.

Stone, R. (1961). *Input-output and national accounts*. Organisation for European Economic Co-operation Paris.

Stone, R. (1962). Multiple classifications in social accounting. *Bulletin of the International Statistical Institute 39*(3), 215–233.

Stone, R., D. G. Champernowne, and J. E. Meade (1942). The precision of national income estimates. *The Review of Economic Studies 9*(2), 111–125.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting 1*, 135–196.

Wei, X. and Y. Yang (2012). Robust forecast combinations. *Journal of Econometrics 166*(2), 224–236.

Zellner, A. and J. Tobias (2000). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting 19*(5), 457–465.