University of Konstanz
Department of Economics

Combining Country-Specific Forecasts when Forecasting Euro Area Macroeconomic Aggregates

Jing Zeng

Working Paper Series
2015-11

Combining country-specific forecasts when forecasting Euro area macroeconomic aggregates *

Jing Zeng[†]

May 13, 2015

Abstract

European Monetary Union (EMU) member countries' forecasts are often combined to obtain the forecasts of the Euro area macroeconomic aggregate variables. The aggregation weights which are used to produce the aggregates are often considered as combination weights. This paper investigates whether using different combination weights instead of the usual aggregation weights can help to provide more accurate forecasts. In this context, we examine the performance of equal weights, the least squares estimators of the weights, the combination method recently proposed by Hyndman et al. (2011) and the weights suggested by shrinkage methods. We find that some variables like real GDP and GDP deflator can be forecasted more precisely by using flexible combination weights. Furthermore, combining only forecasts of the three largest European countries helps to improve the forecasting performance. The persistence of the individual data seems to play an important role for the relative performance of the combination.

Keywords: Forecast combination, aggregation, macroeconomic forecasting, hierarchical time series, persistence in data

JEL classification: C22, C43, C53

*I thank the participants of the Doctoral Seminar on Econometrics at the University of Konstanz and the Konstanz-Lancaster Workshop on Finance and Econometrics for helpful comments and suggestions. Financial support by the Deutsche Forschungsgemeinschaft, project number BR 2941/1-2, is gratefully acknowledged.

[†]University of Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: jing.zeng@uni-konstanz.de

1 Introduction

Forecasting Euro area macroeconomic aggregates is of interest to economic agents and policy makers at all levels. One important issue in this context is whether the forecasts of the aggregates should be constructed by using only the past information of the aggregates or whether disaggregate forecasts for different European Monetary Union (EMU) member countries are obtained first and then combined. The theoretical literature concerning the effect of contemporaneous aggregation on the forecasting performance (Tiao and Guttman (1980), Wei and Abraham (1981) and Lütkepohl (1984a,b)) derives the important result that when the true data generating process (DGP) is known, then aggregating disaggregate forecasts leads to lower mean squared forecast errors (MSFE) than using only aggregate information in forecasting the aggregate and can therefore increase the forecasting precision. Surveys on forecasting aggregate time series data can be found in Granger (1990) and Lütkepohl (2010).

However, in practice, the true DGP will be unknown, thus the forecasting accuracy depends on the empirical data under analysis. A number of studies document this in the literature. Marcellino et al. (2003) compare these two approaches for forecasting four Euro area macroeconomic aggregates over the period from 1982 to 1997 and find that the most accurate forecasts are produced by combining 11 country-specific forecasts. The study of Hubrich (2005) focuses on investigating whether aggregating forecasts of subcomponents (unprocessed and processed food, industrial good, energy and services prices) of the Harmonized Index of Consumer Prices (HICP) in the Euro area can improve the forecasting accuracy. She finds that in the presence of some shocks (e.g. the oil price shock), the forecast errors of the disaggregate subcomponents tend to have the same sign and cannot cancel out. Consequently, combining the disaggregate forecasts leads to increased bias when predicting the year-on-year inflation twelve months ahead. However, if "core" inflation, which does not include unprocessed food and energy prices, is taken into account, a better forecasting performance is observed when combining the subcomponent forecasts. This may be due to the fact that the subcomponents included in the "core" inflation are less affected by the shocks. Zellner and Tobias (2000) investigate the effects of aggregation and disaggregation in forecasting the median growth rate of 18 industrialized countries and provide evidence that improved forecasting results can be achieved by disaggregation, however, the aggregate variable of interest should be included in the single disaggregate equation. Further developments in this area suggest to use disaggregate variables or information extracted from disaggregate variables in forecasting the aggregates, see for example Hendry and Hubrich (2006, 2011) and Brüggemann and Lütkepohl (2013).

This paper focuses on combining country-specific forecasts when forecasting Euro area macroeconomic aggregates. As noted in Hubrich (2005), the aggregation of disaggregate forecasts can be considered as a way of forecast combination. Since the first work of Bates and Granger (1969) which motivates the idea of combining different forecasts, enormous developments can be observed in the literature concerning this topic, see for example Clemen (1989), Granger (1989), Hendry and Clements (2004), Wallis (2011) and Hsiao and Wan (2014). Reasons in favour of the combination of forecasts are listed in these papers. Generally, by combining multiple forecasts of the same variable, underlying information sets available or observable to different forecasters or organizations are pooled as well. Thus forecast combination can be viewed as a kind of information combination. Moreover, combining forecasts generated by different forecast models and data can reduce the misspecification bias and measurement errors and thus lead to more robust forecasts. Especially if the individual forecasts are differently biased, then combining them may mitigate the forecast errors.

Facing the forecast combination problem, a key issue is how to define the combination weights. The weights can be chosen fixed (for example, equal weights) or estimated (for example, based on their historical forecast performance). Timmermann (2006) reviews various forecast combination methods theoretically and Stock and Watson (2004) compare the empirical performance of these methods. However, it should be noticed that there is an important difference between the normal forecast combination problem discussed in the literature and the forecast combination topic considered in this work. In the literature, all papers concerning this topic aim at combining various forecasts of the same variable. In our work, the individual forecasts used for combination are country-specific forecasts. Thus they are forecasts of different disaggregate components. For example, the aggregate Euro area GDP growth can be forecasted by combining the forecasts of the GDP growth of the member countries. Due to this reason combination methods based on evaluating the relative forecast performance of the single forecast cannot be considered.[1]

When combining the country-specific forecasts, the aggregation weights which are used to construct the aggregate variable are often considered as combination weights. For example, Marcellino et al. (2003) use the fixed 1997 real GDP share weights to combine the 11 country-specific disaggregate forecasts. These weights are the same as those used to produce the aggregate. In this paper, we examine the empirical performance of different combination weights when combining the country-specific forecasts for forecasting four Euro area macroeconomic aggregate variables. The first group of combination weights includes equal weights and the least squares (LS) estimators of the weights which are popular in the usual forecast combination literature. The second group of combination weights is based on the original aggregation weights. We apply the novel combination strategy proposed by Athanasopoulos et al. (2009) and Hyndman et al. (2011) and the shrinkage method suggested by Stock and Watson (2004) to derive combination weights that also use information of the original aggregation weights. It should be noticed that one could also try to improve the forecasting performance of the forecast combination by selecting the most accurate forecast model for each countries' time series and then combine them. However, the main purpose of this paper is to examine the impact of different combination weights. Thus we use the same country-specific forecasts (estimated by using the purely univariate autoregressive models) for different combination weights.[2]

[1] It is also possible to forecast the Euro area aggregate variables by using just data from one EMU member country as predictors. For example, Brüggemann et al. (2008) forecast the Euro area variables with German pre-EMU data. Thus, by combining forecasts which use data from different member countries as predictors, estimated combination weights considering the forecast accuracy of each single forecast can be used. This is also examined in our work. However, the results show that this way of forecast combination cannot reduce the mean squared forecast errors. Thus the results are not reported in this work.

[2] Marcellino (2004) compares a large number of linear and nonlinear models for fore-

We evaluate the forecasting performance of different combination methods in forecasting four Euro area macroeconomic aggregate variables: the real GDP, the GDP deflator, the Consumer Price Index (CPI) and the long-term interest rates. Recursive pseudo-out-of-sample forecasting experiments are conducted. We find that more flexible combination weights (e.g. the LS combination weights and the weights suggested by using the shrinkage method of Stock and Watson (2004)) are needed when forecasting the aggregate real GDP and GDP deflator, while using fixed aggregation weights as combination weights beats the forecast of the aggregate variables based on their own past information at all considered horizons for the CPI and long-term interest rates. Furthermore, combining only forecasts of the three largest European countries Germany, France and Italy helps to improve the forecasting performance. An additional finding of this paper is that if the member countries data exhibit different persistence, for example, some countries data are more persistent, while the others are less persistent, then choosing only forecasts of those countries whose data are highly persistent for the combination may help to provide more accurate forecasts.

The remainder of this paper is organized as follows. In Section 2, we briefly discuss the equal weights and the least squares estimators of the weights which are popular in the usual forecast combination literature. Section 3 introduces the novel combination strategy proposed by Athanasopoulos et al. (2009) and Hyndman et al. (2011). The shrinkage method suggested by Stock and Watson (2004) for forecast combination is discussed in Section 3 as well. The empirical results of our forecasting experiments are presented and described in Section 4. Finally, Section 5 summarizes the main results and concludes.

2 Popular forecast combination weights

The choice of combination weights is an important issue when combining forecasts. Simple combination forecasts discussed in Stock and Watson (2004) use equal weights or only the (trimmed) mean of all individual forecasts.

casting aggregate EMU macroeconomic variables and the main result is that for a number of variables the simple autoregressive models perform quite well.

Combination weights which are computed based on evaluating the relative performance of the single forecast are originally proposed by Bates and Granger (1969). Timmermann (2006) provides a thorough overview of combination methods which estimate the combination weights. Swanson and Zeng (2001) examine forecast combination based on a model selection approach. More recently, Hsiao and Wan (2014) introduce an eigenvector approach to determine the combination weights. In this section, the simple equal weights used for combination are discussed at first, since the equal weighting has been found in the literature to have better performance than more complicated combining methods. The second method introduced here estimates the combination weights by the ordinary least squares. Other more complicated estimated combination weights are not analyzed because the finite-sample errors in estimating the weights can affect the combination negatively. Combination weights based on the historical performance of the individual forecasts are not considered because the disaggregate forecasts used for combination are forecasts of different country-specific forecasts.

2.1 Equal weights

The result that using simple mean combination forecast with equal weights outperforms other complicate forecast combinations is confirmed in many studies, see Clemen (1989) for a summary of these papers up to 1989. Stock and Watson (2004) compare different forecast combination methods based on 43 time series of seven developed countries and find that the simple average of many forecasts leads to improvement of forecasting performance. Genre et al. (2013) investigate a number of more sophisticated combination strategies compared to the equal weights to combine the expert forecasts from the European Central Bank Survey of Professional Forecasters. They conclude that only few of these combination schemes outperform the simple average. Smith and Wallis (2009) point out that the reason lies in the estimation errors. When complex forecast combination methods are applied and the aggregation weights need to be estimated, then these estimated combination weights may be unreliable or unstable, such that the expected advantage of combining forecasts disappears.

Although the member countries' variables are aggregated via non-equal weights to produce the aggregates in our work, it is interesting to examine whether using the equal weights to combine the country-specific forecasts is a feasible alternative. Suppose the multi-period h-step-ahead forecasts of the Euro area aggregates and the N individual EMU countries data are denoted by $y_{t+h|t}^{EMU}$ and $x_{i,t+h|t}$ for $i = 1, \ldots, N$, respectively. The simple mean combination forecast can be expressed as in the following equation:

$$\hat{y}_{t+h|t}^{EMU} = \frac{1}{N}\sum_{i=1}^{N}\hat{x}_{i,t+h|t} = \frac{1}{N}\hat{x}_{1,t+h|t} + \ldots + \frac{1}{N}\hat{x}_{N,t+h|t},$$

where $\hat{y}_{t+h|t}^{EMU}$ and $\hat{x}_{i,t+h|t}$ denote the aggregate and disaggregate forecasts based on models with estimated parameters.

2.2 Least squares weights

Granger and Ramanathan (1984) suggest to estimate the forecast combination weights by using the ordinary least squares (LS) method. Three versions of regression models are considered:

$$\begin{aligned}
y_{\tau+h}^{EMU,(1)} &= w_{0,h}^{LS1} + w_{1,h}^{LS1}\hat{x}_{1,\tau+h|\tau} + \ldots + w_{N,h}^{LS1}\hat{x}_{N,\tau+h|\tau} + \varepsilon_{\tau+h}, && (2.1)\\
y_{\tau+h}^{EMU,(2)} &= w_{1,h}^{LS2}\hat{x}_{1,\tau+h|\tau} + \ldots + w_{N,h}^{LS2}\hat{x}_{N,\tau+h|\tau} + \varepsilon_{\tau+h}, && (2.2)\\
y_{\tau+h}^{EMU,(3)} &= w_{1,h}^{LS3}\hat{x}_{1,\tau+h|\tau} + \ldots + w_{N,h}^{LS3}\hat{x}_{N,\tau+h|\tau} + \varepsilon_{\tau+h}, &&\\
s.t. \quad & \sum_{i=1}^{N} w_{i,h}^{LS3} = 1. && (2.3)
\end{aligned}$$

The regression model in (2.1) allows an intercept $w_{0,h}^{LS1}$. If there exists bias in the single disaggregate forecasts, then the bias can be captured through this intercept term such that the forecast for the aggregate may still be unbiased. In contrast, the regression models (2.2) and (2.3) are more restrictive. They assume that the disaggregate forecasts are unbiased. Moreover, by restricting the sum of the weights equal to one, regression (2.3) ensures that the forecast for the aggregate is unbiased given the assumption that the disaggregate forecasts are unbiased.

If we want to forecast an aggregate variable based on information until the period t, by using this method, we start with the first estimation period from the beginning of the sample until the period τ_0 and obtain the first set of estimated disaggregate forecasts $\hat{x}_{1,\tau_0+h|\tau_0}, \ldots, \hat{x}_{N,\tau_0+h|\tau_0}$. This estimation period is then increased by one observation and the second set of estimated disaggregate forecasts can be obtained. We continue this process until the period $t-h$. At the end, the observed values of this aggregate variable over the period $\tau = \tau_0, \ldots, t-h$ are regressed on the estimated disaggregate forecasts to obtain the estimated combination weights. The combination weights in (2.1) and (2.2) can be estimated by using the usual LS method and the combination weights in (2.3) are estimated through the restricted LS method (see e.g. Greene and Seaks (1991)). The estimated combination weights are then used to compute the forecast for the aggregate $\hat{y}^{EMU}_{t+h|t}$ as:

$$\begin{aligned}
\hat{y}^{EMU,(1)}_{t+h|t} &= \hat{w}^{LS1}_{0,h,t} + \hat{w}^{LS1}_{1,h,t}\hat{x}_{1,t+h|t} + \ldots + \hat{w}^{LS1}_{N,h,t}\hat{x}_{N,t+h|t}, \\
\hat{y}^{EMU,(2)}_{t+h|t} &= \hat{w}^{LS2}_{1,h,t}\hat{x}_{1,t+h|t} + \ldots + \hat{w}^{LS2}_{N,h,t}\hat{x}_{N,t+h|t}, \\
\hat{y}^{EMU,(3)}_{t+h|t} &= \hat{w}^{LS3}_{1,h,t}\hat{x}_{1,t+h|t} + \ldots + \hat{w}^{LS3}_{N,h,t}\hat{x}_{N,t+h|t},
\end{aligned}$$

respectively.

3 Forecast combination weights based on the aggregation weights

Aggregate variables are typically weighted sums of disaggregate components. In our context, the weights used to construct the Euro area macroeconomic aggregates often reflect the relative size of the country within the EMU (measured as the GDP shares). Clearly, the weights used in aggregation may also be considered for forecast combination (e.g. Marcellino et al. (2003) and Lütkepohl (2010)). In this section, the aggregation weights will be briefly described first. Then two variants of combination weights based on the aggregation weights are introduced.

3.1 Aggregation weights

The Euro area aggregates y_t^{EMU} can be defined as the weighted sum of individual EMU countries data $x_t = (x_{1,t}, \ldots, x_{N,t})'$:

$$y_t^{EMU} = w'x_t = w_1 x_{1,t} + \ldots + w_N x_{N,t}. \tag{3.1}$$

$x_{i,t}$ for $i = 1, \ldots, N$ can be levels data or log-levels data or growth rates from each country. The aggregation weights $w_1, \ldots, w_N$ are often fixed and time-invariant. Moreover, they satisfy $w_i > 0$ and $\sum w_i = 1$. The most widely used method for constructing the Euro area data is provided in Fagan et al. (2001, 2005). This aggregation is performed on log-levels data using fixed weights. The Area Wide Model (AWM) database which provides a wide range of quarterly Euro area macroeconomic time series data available from 1970 is based on this method.

In the literature, the aggregation weights can also be time-varying. Beyer et al. (2001) suggest to aggregate the growth rates of the variables by using time-varying weights based on the real GDP share of the previous period. Similarly to this idea, Anderson et al. (2011) propose to use sliding weights for aggregation which measure the time-varying distance of the periphery countries from the core countries in terms of monetary integration. The choice of different aggregation methods may have enormous impact on the subsequent empirical analysis, see e.g. Bosker (2006). The impact of time-varying weights on forecasting contemporaneous aggregates is analyzed in Brüggemann and Lütkepohl (2013). In our work, time-varying aggregation weights are not considered because the combination weights discussed in Section 3 assume fixed aggregation weights.

Suppose the multi-period ahead forecasts of the Euro area aggregates and the individual countries data are denoted by $y_{t+h|t}^{EMU}$ and $x_{i,t+h|t}$ for $i = 1, \ldots, N$. A natural way of combining the country-specific forecasts to obtain the forecasts of the aggregates is to use the vector of aggregation weights $w = (w_1, \ldots, w_N)'$ as in equation (3.1) to obtain:

$$\hat{y}_{t+h|t}^{EMU} = w_1 \hat{x}_{1,t+h|t} + \ldots + w_N \hat{x}_{N,t+h|t}.$$

Again, $\hat{y}_{t+h|t}^{EMU}$ and $\hat{x}_{i,t+h|t}$ denote the aggregate and disaggregate forecasts based on models with estimated parameters. In the following subsections, two alternative combination methods will be introduced which do not use the aggregation weights $w_1, \ldots, w_N$ as forecast combination weights directly. However, they are defined as functions of $w_1, \ldots, w_N$ from Equation (3.1).

3.2 Optimal forecast combination for hierarchical time series

Hierarchical time series are hierarchically structured and can be aggregated at different levels in groups. An example presented in Athanasopoulos et al. (2009) and Hyndman et al. (2011) are the domestic tourism data which are disaggregated by purpose of travel or by geographical region. Since the EMU data are also disaggregated at the geographical country level, the suggested approach may be useful in forecasting European aggregate time series.

The commonly applied methods in forecasting hierarchical data are the bottom-up approach which starts with the forecasts for the bottom level and then aggregates those forecasts to produce forecasts for higher hierarchies and the top-down approach which conversely disaggregates the top level forecasts down according to some rules[3]. The comparative forecasting performance of these two approaches is discussed in Grunfeld and Griliches (1960), Shlifer and Wolff (1979), Schwarzkopf et al. (1988) and Fliedner (1999). To summarize, the bottom-up method better captures the structure of the hierarchy due to accumulation from the bottom level. However, if the forecasts at the lowest level are biased or incomplete because of missing data, then this will lead to larger bias in forecasting at the higher levels. The top-down approach provides more accurate forecasts for the top level forecasts, but it cannot take the individual characteristics of time series at each level into account.

The new combination approach proposed in Athanasopoulos et al. (2009) and Hyndman et al. (2011) considers smaller hierarchies with less than four levels, while hierarchies with millions of time series at the lowest level are discussed in Hyndman et al. (2014). Starting point of this approach is to

[3]Athanasopoulos et al. (2009) discuss two versions of the top-down approach. One approach disaggregates the top level forecasts to produce the lower level forecasts based on the historical proportions of the lower level series relative to the top aggregate. The other one is based on the forecasted propotions.

forecast all time series at all (aggregate and disaggregate) levels first. The obtained forecasts are referred to as base forecasts. Then the generalized least squares estimation is used to produce the minimum variance unbiased combination of these base forecasts which also satisfies the hierarchy. More detailed description of the derivation and proof of this method can be found in Hyndman et al. (2011).

The central idea of this approach is to revise the base forecasts at hand by exploiting the hierarchical structure. No restrictions on the computation of the base forecasts are imposed. In principle, the base forecasts may be obtained by choosing from possible forecasting methods discussed in the literature and with all kinds of information. This may include models that use information from aggregate variables or from disaggregate variables or from both. Other exogenous variables may be included in the information set as well. If the base forecasts are unbiased, this approach simply provides revised forecasts which are also unbiased and have the smallest variance among all possible combinations that satisfy the hierarchical structure. So far, this novel method has mostly been used in the context of tourism forecasting. Athanasopoulos et al. (2009) and Hyndman et al. (2011) find that more accurate forecasts can be obtained by using this kind of forecast combination in forecasting Australian domestic tourism demand, while Andrawis et al. (2011) apply this method to combine short-term (monthly) and long-term (yearly) forecasts and analyze the tourism demand for Egypt. Capistran et al. (2010) select at first the best forecasting models for each subcomponent of the Mexican Consumer Price Index and then combine these forecasts by using this method. The results indicate that this approach yield better forecasts than just combining the forecasts from the bottom level by using the corresponding weights.

According to the characteristics of the Euro area data, we consider a hierarchical structure with only two levels in this work. The top level contains the aggregate series y_t^{EMU} and the bottom level contains the N disaggregate data $x_{1,t}, \ldots, x_{N,t}$. For each time observation $t = 1, \ldots, T$, the hierarchical structure of the aggregate and disaggregate data is shown in Figure 1. It should be noticed that in the applications in the literature discussed above, more than three levels are considered. In our context, it is also possible to

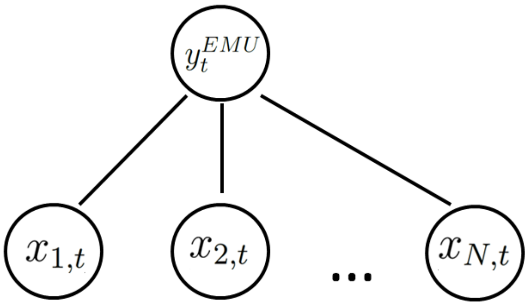


Figure 1: The hierarchical structure of the Euro area aggregates and disaggregate data

include more levels in the hierarchy. For example, when combining the real GDP growth of all EMU member countries to obtain the forecast of the Euro area aggregate GDP growth, subcomponents of real GDP (private and government final consumption, gross fixed capital formation and net exports) from each country can also be considered. However, the main purpose of this work is to investigate the effect of different combination weights when combining country-specific forecasts. If more information of the subcomponents from each country is used for this approach, we could not identify whether the use of more information or the use of these combination weights is the reason for possible improvements of forecasting performance.

Let Y_t denote the vector including all the aggregate and disaggregate time series at time t: $Y_t = (y_t^{EMU}, x_{1,t}, \ldots, x_{N,t})'$ and define a "summing" matrix S of dimension $(N+1) \times N$ which has the following form:

$$S = \begin{pmatrix} w' \\ I_N \end{pmatrix},$$

where $w = (w_1, \ldots, w_N)'$ is the vector of aggregation weights defined in equation (3.1). Now we write

$$Y_t = Sx_t,$$

or to be more explicit

$$\begin{pmatrix} y_t^{EMU} \\ x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{N,t} \end{pmatrix} = \begin{pmatrix} w_1 & w_2 & \dots & w_N \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ \vdots \\ x_{N,t} \end{pmatrix}.$$

In this approach, the h-step-ahead base forecasts for the aggregate and disaggregate time series based on the information set available up to time t are computed first:

$$\hat{Y}_{t+h|t} = (\hat{y}_{t+h|t}^{EMU}, \hat{x}_{1,t+h|t}, \dots, \hat{x}_{N,t+h|t})'.$$

Denoting the revised forecasts by $\tilde{Y}_{t+h|t}$, Athanasopoulos et al. (2009) and Hyndman et al. (2011) show that the revised forecasts $\tilde{Y}_{t+h|t}$ are given by

$$\tilde{Y}_{t+h|t} = S(S'S)^{-1}S'\hat{Y}_{t+h|t}. \tag{3.2}$$

Obviously, the revised forecasts $\tilde{Y}_{t+h|t}$ in equation (3.2) are linear combinations of the base forecasts. Notice that the aggregation weights used for hierarchical time series in Athanasopoulos et al. (2009) and Hyndman et al. (2011) are $w_1 = \ldots = w_N = 1$. However, in our forecasting analysis with the Euro area data, the aggregation weights are not equal, but add up to one ($\sum w_i = 1$).[4] Therefore, according to equation (3.2), the revised forecast for the Euro area aggregates $\tilde{y}_{t+h|t}^{EMU}$ is a linear combination of the base forecasts of both aggregate and disaggregate variables $\hat{y}_{t+h|t}^{EMU}, \hat{x}_{1,t+h|t}, \dots, \hat{x}_{N,t+h|t}$. The combination weights $w_1^h, \dots, w_{N+1}^h$ implied by the first row of $S(S'S)^{-1}S'$ are functions of $w_1, \dots, w_N$:

$$\tilde{y}_{t+h|t}^{EMU} = w_1^h \hat{y}_{1,t+h|t}^{EMU} + w_2^h \hat{x}_{1,t+h|t} + \ldots + w_{N+1}^h \hat{x}_{N,t+h|t}.$$

Moreover, by considering all information available within a hierarchy, this method provides adjustments not only for the forecasts of the aggregate variables but also for the single disaggregate forecast.

[4] Using nonequal weighs still provide the minimum variance unbiased combination, given that the base forecasts are unbiased, since the derivation and proof of this method are independent of the form of the S matrix.

One drawback of these combination weights is that the revision of the forecasts in (3.2) does not take the actual data into account. Only the hierarchy structure matters. Suppose two different empirical variables which have the same hierarchy, then this method will provide the same combination scheme. Furthermore, this method is inappropriate for the case where time-varying weights are used in aggregation.

3.3 Shrinkage method for forecast combination

Diebold and Pauly (1990) suggest to incorporate prior information on the weights into the estimation of forecast combination weights by using Bayesian shrinkage techniques. Based on the g-prior model of Zellner (1986), they derive that the vector of the posterior weights w^S has the following form:

$$w^S = w^0 + \frac{1}{1+g}(w^{LS} - w^0). \tag{3.3}$$

w^0 represents the vector of the prior weights and w^{LS} contains the LS estimators of the combination weights which are discussed in Section 2.2. The parameter g which controls the amount of shrinkage can be estimated by using empirical Bayes methods (see Diebold and Pauly (1990)). It is easy to see that the combination weights in w^S are a linear combination of the prior weights and the LS estimators of the combination weights. A larger value of g implies shrinkage towards the prior information and a smaller g leads to less shrinkage.

Diebold and Pauly (1990) use equal weights as the prior weights. The same choice can be found in Stock and Watson (2004). Moreover, for each weight in w^S, Stock and Watson (2004) simplify the equation (3.3) as:

$$w_i^S = \lambda w_i^{LS} + (1-\lambda)1/N. \tag{3.4}$$

The parameter λ is determined by:

$$\lambda = \max(0, 1 - \kappa N/(t - h - \tau_0 - N)), \tag{3.5}$$

where N denotes the number of forecasting models to be combined. Data available up to the period t are used to estimate the LS weights and h is the forecasting horizon. τ_0 is defined as in Section 2.2. κ is the strength of the shrinkage and Stock and Watson (2004) evaluate forecasts for $\kappa = 0.25, 0.5$ and 1. Larger values of κ correspond smaller values of λ and hence more shrinkage towards equal weighs $1/N$. The value of the parameter λ is restricted to be nonnegative and depends on N and the number of observations used for estimating the LS weights (t). If t increases faster than N, in the limit the LS weights will obtain a weights of unity. The weights in equation (3.4) are quite often discussed and used in the forecast combination literature (see for example Timmermann (2006) and Wright (2008)).

We modify the idea of using equal weights as prior weights by considering the aggregation weights of equation (3.1) in (3.4):

$$w_i^S = \lambda w_i^{LS} + (1 - \lambda) w_i. \tag{3.6}$$

According to equation (3.6), the combination procedure based on shrinkage methods provides a convex combination of the LS weights and the original aggregation weights. Since the parameter in (3.5) depends only on the relative proportion of N and t, by using the aggregation weights instead of the equal weights as prior information, λ of equation (3.6) is also defined as in (3.5). Thus, to compute the forecast of the aggregate variables, the disaggregate forecasts can be combined with the weights in (3.6):

$$\hat{y}_{t+h|t}^{EMU} = w_{1,t}^S \hat{x}_{1,t+h|t} + \ldots + w_{N,t}^S \hat{x}_{N,t+h|t}.$$

Notice that the weights given in (3.6) depend on the sample size considered for the estimation of LS weights, thus w^S will change if different recursive samples are used in the forecasting exercise.

4 Empirical analysis

Since the introduction of the Euro in 1999, a large number of empirical papers have acknowledged the importance of analyzing and forecasting the Euro area macroeconomic aggregates. Marcellino et al. (2003) point out that "political and business decisions increasingly depend on aggregate European real economic activity". For example, Marcellino (2004) focuses on forecasting the EMU macroeconomic variables and evaluating the performance of a large number of forecasting methods. Banerjee et al. (2005) investigate the role of a set of variables as leading indicators for Euro area inflation and GDP growth. Hubrich (2005) examines whether aggregating forecasts of Harmonized Index of Consumer Price subcomponents lead to more accurate forecasts of Euro area inflation. Our empirical analysis focuses on how the forecasting accuracy of the Euro area aggregate variables are affected by choosing different combination weights when combining the country-specific forecasts.

4.1 Data

In the empirical study, four macroeconomic variables, the real GDP (YER), the consumer price index (CPI), the GDP deflator (YED) and the long-term interest rate (LTN) are considered.[5] Forecasts of Euro area inflation and GDP growth are of central interest for the European Central Bank and policy makers.

The aggregate time series data are taken from the Area Wide Model (AWM) database over the period from 1970Q1 to 2011Q4. The Euro Area Business Cycle Network[6] reports a wide range of quarterly Euro Area macroeconomic time series. Nowadays the AWM database is widely used for empirical studies on the Euro Area economy (see e.g. Banerjee et al. (2005) and Küster and Wieland (2010)). The aggregate data are constructed according to Fagan et al. (2001, 2005). For example, the logarithm of the aggregate GDP for the Euro area is a weighted average of the logarithms of the member countries' GDP, where fixed weights based on the real GDP shares for 1995 are used. Since in Fagan et al. (2001) only 11 aggregation weights (without weight for Greece) are given, we calculate the weights for 12 European countries by using the same method described in Fagan et al. (2001) and report the results in the column with w^0 in Table 1. Disaggregate data from 12

[5]The exchange rate against the US-Dollar and the short-term interest rates are not considered in this work, since with the introduction of the Euro in 1999 all the Euro zone countries have the same data for these two variables respectively. When applying the LS combination weights, the problem of singular matrix occurs due to the same data.

[6]The corresponding website is http://www.eabcn.org.

Table 1: AWM weights updated for 12 countries (w^0) and the combination weights derived from the approach for hierarchical time series (w^h)

countries	w^0	w^h	countries	w^0	w^h
Austria	0.030	0.025	Ireland	0.012	0.010
Belgium	0.037	0.031	Italy	0.191	0.161
Finland	0.017	0.014	Luxembourg	0.003	0.002
France	0.213	0.179	Netherlands	0.061	0.051
Germany	0.303	0.254	Portugal	0.018	0.016
Greece	0.018	0.015	Spain	0.097	0.082

Note: w^h is based on the method suggested by Athanasopoulos et al. (2009) and Hyndman et al. (2011). This method assigns a weight to the forecast of the aggregate variable in the forecast combination. In our example this weight is 0.160.

EMU countries are considered here[7]. The quarterly data are taken from the OECD main economic indicators and range from 1970Q1 to 2011Q4.

First, all the time series are seasonally adjusted. Second, if the data on real GDP, CPI and GDP deflator have different base years, they are rebased such that all series refer to the same base year 2005. Third, augmented unit-root tests are applied to check the integration properties of these time series. Since all data are integrated of order one, the first differences of logarithms are taken for the variables YER, CPI and YED, while for the variable LTN only the first differences of levels are used. At last, since not all variables are available for the whole sample period, we use the expectation-maximization (EM) method suggested in Stock and Watson (002a) to interpolate missing values. Time series with missing values are listed in Table 2.

4.2 Forecasting details

The main purpose of this paper is in investigating whether using different combination weights to combine country-specific forecasts can produce more

[7]12 European countries considered in this work are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain.

Table 2: List of variables with missing observations for some countries

Countries	Series	Periods with missing observations
Greece	LTN	1970Q4-1994Q4
Ireland	YER	1970Q1-1989Q4
	CPI	1970Q1-1975Q4
	YED	1970Q1-1989Q4
	LTN	1970Q1-1989Q4
Luxembourg	LTN	1970Q1-1993Q3

accurate forecasts for the aggregates, compared to the use of the aggregation weights as combination weights. Thus, rather than finding the best performing forecast model for each country, we focus on the simple univariate autoregressive models with two lags for forecasting the countries data[8] to obtain the disaggregate forecasts. The AR(2) model used is:

$$x_{i,t+h} = c_{i,h} + \alpha_{i,1} x_{i,t} + \alpha_{i,2} x_{i,t-1} + \varepsilon_{i,t+h}. \tag{4.1}$$

For each variable, we estimate equation (4.1) for each of the 12 countries data by using ordinary least squares (OLS) in a recursive manner. The first estimation sample starts in 1970Q1 and ends in 2002Q4. With the estimated coefficients, forecasts for $h = 1, 2$ and 4 steps ahead for each country are computed. Thus 12 country-specific forecasts are obtained. For the second estimation sample from 1970Q1 to 2003Q1, the same exercise is repeated. Since data are available until 2011Q4, the last estimation sample ends in 2010Q4. Thus, we obtain 33 pseudo-out-of-sample forecasts for each horizon h. For the forecasting performance evaluation, we use the mean squared forecast error (MSFE), where the forecast error is defined as the difference of the true value and the estimated forecasted value of the aggregate variable. Because of its good forecasting performance, the univariate autoregressive models with four lags of the aggregate variable is considered as our bench-

[8] A robustness check for different lag length choice is also undertaken. Using the Akaike information criteria for selecting the lag length does not change the main findings of this paper.

mark. We compare the MSFE of different combination methods against the MSFE obtained from the benchmark. Combination methods which perform better than the benchmark should have smaller relative mean squared forecast errors (relative MSFEs) than one.

4.3 Results of forecasting comparison

The relative MSFEs of different combination methods for each of the considered variable and each forecasting horizon are presented in Table 3. We report results for forecast combinations using the AWM aggregation weights (w^0), the equal weights (w^e), the three variants of the LS weights (w^{LS1}, w^{LS2} and w^{LS3}), the weights derived from the approach for hierarchical time series (w^h) and the weights computed by using the shrinkage method (w^{S1}, w^{S2} and w^{S3}). In line with Stock and Watson (2004), for the shrinkage method, the LS estimators of the weights w^{LS} used in (3.6) are the estimated coefficients of the regression model without an intercept, as in (2.2). Thus, the combination weights are computed by using the shrinkage method for three different values of κ in (3.5): $\kappa = 0.25, 0.5$ and 1, denoted as w^{S1}, w^{S2} and w^{S3}, respectively. When estimating the LS weights in regression models (2.1), (2.2) and (2.3) and in equation (3.6), τ_0 is set as 1989Q4.

The results in Table 3 indicate that for the variables real GDP (YER) and GDP deflator (YED), using the original AWM aggregation weights as forecast combination weights leads to larger MSFEs relative to the AR(4) benchmark for the aggregate. Especially for the GDP deflator, the relative MSFEs are substantially larger than one at all forecasting horizons. This indicates that forecasting the aggregate with a simple AR(4) model performs better than aggregating the country-specific forecasts by using the original aggregation weights. Furthermore, we find that using the combination weights derived from the approach for hierarchical time series does not change the relative weights of the disaggregate forecasts. It gives only part of the total weights to the forecast of the aggregate variable. These weights are reported in Table 1 together with the AWM aggregation weights. We see that the forecast of the aggregate variable obtains the weight of 0.16 when combining all the base forecasts. Thus, the weight assigned to all the country-specific forecast

Table 3: Relative MSFEs of different combination methods using 12 country-specific forecasts

	w^0	w^e	w^{LS1}	w^{LS2}	w^{LS3}	w^h	w^{S1}	w^{S2}	w^{S3}
	YER								
$h = 1$	1.017	1.092	0.845	0.905	0.900	1.009	0.899	0.893	0.887
$h = 2$	0.973	1.000	1.120	1.110	1.110	0.975	1.101	1.092	1.076
$h = 4$	1.021	1.060	1.071	1.096	1.085	1.017	1.087	1.080	1.065
	YED								
$h = 1$	1.251	1.482	0.893	0.881	1.059	1.177	0.890	0.902	0.931
$h = 2$	1.385	1.651	1.150	1.200	1.371	1.289	1.196	1.196	1.200
$h = 4$	1.280	1.667	0.979	0.954	1.093	1.218	0.942	0.935	0.934
	CPI								
$h = 1$	0.964	1.109	1.137	1.161	1.129	0.961	1.131	1.103	1.054
$h = 2$	0.942	1.075	0.904	0.960	0.967	0.948	0.952	0.944	0.931
$h = 4$	0.972	1.104	0.941	1.263	1.162	0.972	1.234	1.206	1.155
	LTN								
$h = 1$	0.980	0.953	1.152	1.140	1.128	0.977	1.125	1.111	1.084
$h = 2$	0.960	0.950	1.095	0.957	0.951	0.965	0.955	0.953	0.950
$h = 4$	0.986	0.982	0.997	0.987	0.996	0.987	0.982	0.978	0.971

Note: The table shows relative MSFEs of different combination approaches. The benchmark is the univariate autoregressive model (AR(4) with an intercept) for the aggregate variable. Univariate AR(2) models with an intercept are used for forecasting the disaggregate variables. The relative MSFEs are computed for out-of-sample recursive forecasts over the period from 2003Q1 to 2011Q4. w^0 refers to the AWM aggregation weights. The equal weights of 1/12 are denoted as w^e. The three variants of the LS weights are w^{LS1}, w^{LS2} and w^{LS3}. The weights derived from the approach for hierarchical time series are denoted as w^h. The combination weights computed by using the shrinkage method for three different values of κ ($\kappa = 0.25, 0.5$ and 1) are denoted as w^{S1}, w^{S2} and w^{S3}, respectively.

is 0.84. For instance, the weight of 0.179 given to France relative to 0.84 results in exactly the AWM weight of 0.213 for France. Due to this reason, the hierarchical time series combination methods should have similar forecast performance as using the AWM weights as combination weights. This is confirmed by comparing the relative MSFEs of these two combination methods in Table 3.

We note that for real GDP and GDP deflator, using more flexible combination weights as in the LS method or the shrinkage method is a useful strategy at least for some forecasting horizons. For YER at $h = 1$, using the LS estimators as combination weights provide more accurate forecasts compared to the benchmark, while for the variable YED, sizable gains in forecasting precision can be observed by using w^{LS1} and w^{LS2}. We also find that the shrinkage method leads to somewhat more precise forecasts for the GDP growth at $h = 1$ and the GDP deflator at $h = 1$ and $h = 4$.

Interestingly, using the AWM aggregation weights w^0 as combination weights when combining the country-specific can beat the benchmark at all considered horizons for the variables CPI and LTN. For these two variables, the relative MSFEs in the column with w^0 are smaller than one. Since using w^h should have similar performance as w^0, as explained before, we observe the same pattern in the column with w^h. Using the LS combination weights and the weights derived from the shrinkage method help to improve the forecasting accuracy of CPI and LTN for some cases. For example, for CPI, using w^{LS1} shows a better performance at larger horizons of $h = 2$ and $h = 4$ than both the benchmark and the use of the AWM aggregation weights, while for the long-term interest rate, compared to w^0, relatively small gains are observed by using the shrinkage method at $h = 2$ and $h = 4$.

Using the equal weights w^e as combination weights leads to larger relative MSFEs and thus worse performance in forecasting the real GDP, GDP deflator and CPI. This indicates that when combining country-specific forecasts for these variables, different countries should be weighted differently. However, using w^e helps to reduce the MSFEs when forecasting the interest rate, even compared to the use of the AWM aggregation weights w^0. This may be due to the fact that the developments in the long-term interest rates in the EMU member countries do not differ widely so that each single country forecast can be weighted equally when combining.

Because the estimated combination weights w^{LS} and w^S lead to much more precise forecasts for YER at $h = 1$ than using the original aggregation weights w^0 as combination weights, we plot them together with the AWM weights for all member countries in Figure 2. Remember that the AMW weights are defined as real GDP share of each member country. These figures suggest that the estimated combination weights have quite different values than the AWM weights. For instance for the variable YER at $h = 1$, for Germany, the estimated combination weights w^{LS} and w^S are larger than its aggregation weight of 0.303, except the line with right-pointing triangle which corresponds to the estimated weights from the regression model with an intercept (w^{LS1})[9]. This means that in the forecast combination, the forecast of Germany has a much larger weight than 0.303. This is also true for the forecasts of Belgium, Finland, France, Greece and Italy which receive larger weights when combining according to the LS method and the shrinkage method, while other small countries like Ireland and Luxembourg obtain very small weights for the combination.

Taking a closer look at the estimated weights plotted in Figure 2 for the real GDP, we observe a strong short rise or drop in the estimated weights around 2010 in some countries (for example Austria, Germany and Italy). The reason for this rapid change lies in the global financial crisis of 2007-2008 which has an enormous negative impact on the real economy of all member countries from 2009 onwards. Consequently, using data until 2009, the estimated weights around 2010 show some sudden changes. If only the sample size from 1970Q1 to 2008Q4 (data before the financial crisis) is taken into account, we find that the advantage of using estimated weights(w^{LS} and w^S) for the forecast combination disappears.[10] This indicates that the LS method and the shrinkage method are more flexible to adapt data in the crisis period.

For the variable YED at $h = 1$ and $h = 4$, using the estimated combination weights w^{LS} and w^S can improve the forecast accuracy compared to the benchmark and the use of AWM weights. Consequently, we also plot the weights in Figure 3 and 4. The solid horizontal lines refer again to the fixed

[9]Due to the intercept, the estimated coefficients in equation (2.1) are quite different to those in (2.2) and (2.3).

[10]Detailed results for this are not reported in this paper, but are available on request.

AWM weights for each country. First of all, we observe that the estimated weights for the GDP deflator are not seriously affected by the financial crisis. By considering the sample size before the financial crisis, the use of estimated combination weights can still provide more accurate forecasts. This could be due to the fact that the data generating processes for Euro area prices have changed less than those for real GDP during the crisis. Different to the plots for the real GDP, we find that for both $h = 1$ and $h = 4$, some of the larger countries in EMU like Germany and Italy receive smaller weights in comparison to AWM weights when combining the forecasts of the GDP deflator from each country, sometimes even negative weights, while other countries like France (at $h = 1$) and Portugal (at $h = 1$ and $h = 4$) obtain large combination weights (larger than their original aggregation weights). This probably reflects the large heterogeneity of inflation rates in the Euro area member countries. Beck et al. (2009) point out that both before and after the introduction of the Euro in 1999, large and persistent differences in inflation rates of EMU member countries are observed. Country-specific factors caused by fiscal policy and special economic factors like production structures or labour market institutions can lead to divergence of inflation rates. Thus, weighting the country-specific inflation rates differently than implied by the AMW weights helps to improve the forecasting performance.

4.4 Combining forecasts of three largest countries in Euro area

In the forecasting exercise reported in Section 4.3, 12 country-specific forecasts are considered for the combination. In this section, only the forecasts of the three largest EMU countries Germany, France and Italy are combined. These three countries are sometimes regarded as the core economies in the Euro area. Clearly, these countries play a major role in constructing the macroeconomic aggregates of the AWM data. Their total GDP share corresponds to about 70%. Thus, we investigate whether it is beneficial to only consider aggregating the forecasting of these three largest EMU economies and therefore set the combination weights of the remaining countries to zero.

Based on the ordering of the countries data used in this work we define the new weighting vector:

$$w^3 = (0, 0, 0, 0.213, 0.303, 0, 0, 0.191, 0, 0, 0, 0)'. \tag{4.2}$$

The weights in w^3 are then used as forecast combination weights. As in Section 4.3, we check the combination of the forecasts of Germany, France and Italy with equal weights using $w^e = 1/3$. The weights estimated by the LS method and the shrinkage method as well as the weights derived from the approach for hierarchial data are also considered. Table 4 presents the results of relative MSFEs.

When forecasting the aggregate GDP growth by combining the forecasts of real GDP growth of Germany, France and Italy, using more flexible combination weights estimated by the shrinkage method (w^{S1}, w^{S2} and w^{S3}) is beneficial to obtain more accurate forecasts. Sizable smaller relative MSFEs are observed in Table 4 for the variable YER at all forecasting horizons, compared to the benchmark. Remember that in Table 3, when combining all 12 country-specific forecasts, using the weights derived from the shrinkage method performs better than the benchmark only at $h = 1$. Using the LS method based on the data of these three countries leads to an improvement in forecast accuracy for the short horizon $h = 1$. For the horizon $h = 2$, smaller gains are obtained by using the weights w^{LS2} and w^{LS3}. Moreover, we observe some gains in forecast precision by using w^3 defined in (4.2) for the horizons $h = 2$ and $h = 4$. Especially at $h = 4$, the smallest relative MSFE is obtained with w^3. The weights derived from the approach for hierarchical data have again quite similar results as w^3.

For the variable GDP deflator, combining the German, French and Italian forecasts of YED by using the fixed AWM aggregation weights of 0.303, 0.213 and 0.191 leads to more accurate forecasts at all horizons compared to the benchmark. Using w^h provides slightly smaller relative MSFEs, compared to the use of w^3. The reason is probably that by using w^h, the forecast of the aggregate GDP deflator receives also a weight for the forecast combination. Taking this into account may contribute to the forecast improvement in comparison to only pooling the forecasts of German, France and Italy together. Furthermore, differently to Table 3, the use of LS method and the shrinkage method for estimating the combination weights is not beneficial here.

Table 4: Relative MSFEs of different combination methods using forecasts of Germany, France and Italy

	w^0	w^3	w^e	w^{LS1}	w^{LS2}	w^{LS3}	w^h	w^{S1}	w^{S2}	w^{S3}
	YER									
$h=1$	1.017	1.013	0.971	0.880	0.866	0.864	1.002	0.870	0.874	0.884
$h=2$	0.973	0.955	0.966	1.054	0.956	0.966	0.955	0.955	0.955	0.954
$h=4$	1.021	0.957	1.013	1.043	0.978	1.027	0.960	0.977	0.975	0.973
	YED									
$h=1$	1.251	0.944	1.317	1.001	1.056	1.069	0.934	1.033	1.012	0.975
$h=2$	1.385	0.927	1.440	1.101	1.139	1.176	0.921	1.114	1.090	1.048
$h=4$	1.280	0.908	1.310	1.101	1.084	1.161	0.905	1.065	1.046	1.013
	CPI									
$h=1$	0.964	1.044	0.941	0.981	1.036	0.954	1.002	1.036	1.036	1.037
$h=2$	0.942	0.909	0.917	0.869	0.940	0.942	0.900	0.938	0.935	0.931
$h=4$	0.972	0.882	0.947	0.832	0.941	0.982	0.887	0.938	0.934	0.928
	LTN									
$h=1$	0.980	0.978	0.992	1.086	1.059	1.000	0.973	1.052	1.045	1.032
$h=2$	0.960	0.953	0.971	1.061	0.988	0.985	0.958	0.986	0.984	0.980
$h=4$	0.986	0.988	0.992	1.059	1.038	1.003	0.989	1.031	1.024	1.012

Note: The table shows relative MSFEs of different combination approaches to combine the forecasts of Germany, France and Italy. The benchmark is the univariate autoregressive model (AR(4) with an intercept) for the aggregate variable. Univariate AR(2) models with an intercept are used for forecasting the disaggregate variables. The relative MSFEs are computed for out-of-sample recursive forecasts over the period from 2003Q1 to 2011Q4. w^0 refers to the AWM aggregation weights. The equal weights of 1/12 are denoted as w^e. The three variants of the LS weights are w^{LS1}, w^{LS2} and w^{LS3}. The weights derived from the approach for hierarchical time series are denoted as w^h. The combination weights computed by using the shrinkage method for three different values of κ ($\kappa = 0.25, 0.5$ and 1) are denoted as w^{S1}, w^{S2} and w^{S3}, respectively.

A possible reason for this observation is that although there is substantial heterogeneity in the inflation rates in the member countries, when taking only the core countries Germany, Italy and France into account, comovements caused by the common factors like monetary policy will dominate so that the AWM weights representing the relative economic importance of the countries can provide more precise forecasts than the estimated combination weights.

For the variable CPI, Table 4 shows that aggregating only the German, French and Italian forecasts of CPI by using w^3 yields better forecasts than combining 12 countries forecasts with w^0 at $h = 2$ and $h = 4$. The combination weights w^h have quite similar performance as w^3. For the horizons $h = 2$ and $h = 4$, the smallest relative MSFEs are obtained by using w^{LS1}. The combination method with the shrinkage method also provides some improvements in forecast accuracy against the benchmark and the combination with w^0, however, it cannot beat the combination with w^3.

For the variable long-term interest rate, using w^3 and w^h leads to slightly smaller relative MSFEs compared to the method of combining the 12 country-specific forecasts by using the AWM aggregation weights w^0 at horizons $h = 1$ and $h = 2$. This result reflects the fact that the long-term interest rates have quite similar developments in all member countries, thus not much information in data from other countries are needed when combining forecasts.

4.5 Persistence of the data

An interesting issue that arises in this empirical analysis is why combining the 12 country-specific forecasts by using the aggregation weights can lead to more precise forecasts of the aggregate CPI and long-term interest rate, while for the variables real GDP and GDP deflator, this approach shows relatively larger MSFEs against the benchmark. To answer this question, we first use the Quandt-Andrews Breakpoint test to detect unknown structural breaks in all time series.[11] There is ample evidence of structural breaks in most time series considered in this work. So the structural change in the data may not provide a reasonable explanation to this question. Next, we check the persistence of the underlying series by estimating univariate autoregressive models with one lag and intercept (AR(1)) for each transformed

[11] The Quandt-Andrews Breakpoint testa are carried out by Eviews 8 for all aggregate and country-specific data. The results are not reported here, but available on request.

time series[12]. To check for possible changes in the persistence of the sample period, we estimate the AR(1) models and report the recursively estimated coefficients together with 95% confidence intervals for the sample periods that end between 1995Q1 and 2011Q4[13]. All the plots of the persistence of the data are summarized in Appendix A.

Figure 5 plots the persistence of the four macroeconomic aggregate variables. Obviously, different variables exhibit different persistent properties. The aggregate CPI shows high persistence with AR parameters around 0.9 and upper limits of 95% confidence intervals ranging between 0.98 and 1.01, followed by YED, while the aggregate YER and LTN are less persistent. Generally, the persistence for the aggregate variable remains quite stable over the time. However, a sudden increase in the estimated AR(1) coefficients for the real GDP growth in 2009 is observed due to the impact of the global financial crisis on the real economy in the Euro area.

The plotted AR(1) coefficients for real GDP growth across countries are shown in Figure 6. For the EMU member countries, there is a clear difference in persistent properties. Some countries like Belgium, France, Greece and Portugal have relative highly persistent real GDP growth series, while much lower persistence is observed for the other countries. Especially the real GDP growth in Germany displays very low persistence (ranging from -0.015 and 0.13) which indicates low serial correlation with its own past. Similar results can be found in Figure 7 for different countries GDP deflator infaltion. However, there are more countries with highly persistent GDP deflator.

Checking the graphs in Figure 8, we find that almost all Euro area countries exhibit high persistence in CPI. It should be emphasised here that the aggregate CPI is highly persistent as well. This result is in line with the findings reported in the econometric literature concerning the inflation persistence in the Euro area, for example Altissimo et al. (2006) and Benati (2008). In contrast, as shown in Figure 9, the countries' LTN are less per-

[12]To obtain stationary time series, the first differences of logarithms are taken for YER, YED and CPI. For the variable LTN the first differences are considered.

[13]The sample for estimation begins always in 1970Q1. We choose 1995Q1 as the end of the first recursive sample because since this date there are no missing values in all considered time series.

sistent and we have also observed fairly low persistence in the aggregate long-term interest rate.

Thus one main conclusion from the persistence check emerges. The aggregate CPI and the CPI in the member countries exhibit all very high persistence, while the aggregate LTN and the countries' LTN show lower persistence. Furthermore, for both variables, there is not much variation in the persistence across countries. This might be the reason why combining the 12 country-specific forecasts by using the aggregation weights can provide more precise forecasts of the aggregate CPI and LTN.[14] In contrast, for the variables real GDP and GDP deflator, much more variation in the persistence parameters is observed. Thus simply combining the country-specific forecasts with the aggregation weights can not help to improve the forecasting performance for these two variables.

Based on the consideration above we select only the forecasts of those countries for combination which have highly persistent data for the variables real GDP and GDP deflator. For real GDP, the forecasts of Belgium, France, Greece and Portugal are combined with their original aggregation weights, while the forecasts of the other countries are not considered for the combination. For the variable GDP deflator, the data in Belgium, France, Greece, Italy, Netherlands, Portugal and Spain are highly persistent and therefore their forecasts are used for combination. The results of forecasting comparison are reported in Table 5. We find that compared to the benchmark and w^0, using selected countries' forecasts for combination can indeed lead to sizable smaller MSFEs and therefore to a better forecasting performance. Therefore, one tentative conjecture from our empirical results is that the underlying persistence of the data seems to play a role for the relative forecasting performance.[15] Since Table 3, 4 and 5 use the same benchmark, by

[14]This finding is not against the results in Hubrich (2005) which forecasts the Euro area HICP by aggregating forecasts of subcomponents. In her paper, if "core" inflation (without the subcomponents unprocessed food and energy prices) is considered, then combining subcomponents forecasts leads to more precise forecasts than forecasting the aggregate by using only its past information.

[15]A simulation study is also conducted. The aim is to investigate whether the persistence of the data can effect the performance of forecast combination. However, the simulation results cannot support our empirical findings in Section 4.5.

Table 5: Relative MSFEs of combination with countries selected according to persistence

	YER		YED	
	w^0	w^{select}	w^0	w^{select}
$h = 1$	1.017	0.875	1.251	0.836
$h = 2$	0.973	0.920	1.385	0.894
$h = 4$	1.021	1.008	1.280	0.871

Note: The table shows relative MSFEs of combination with countries selected according to persistence. The benchmark is the univariate autoregressive model (AR(4) with an intercept) for the aggregate variable. Univariate AR(2) models with an intercept are used for forecasting the disaggregate variables. The relative MSFEs are computed for out-of-sample recursive forecasts over the period from 2003Q1 to 2011Q4. w^0 refers to the AWM aggregation weights. w^{select} refers to the combination method where only countries with highly persistent data are considered for combination.

comparing them we note that for real GDP, using w^{select} does not necessary lead to an improvement in the forecast accuracy compared to using w^{LS2} or the estimated weights suggested by the shrinkage method to combine the forecasts of Germany, France and Italy (Table 4), while for the GDP deflator, using w^{select} beats any other combination methods.

5 Conclusion

This paper investigates whether using different combination weights when combining the country-specific forecasts is beneficial in forecasting Euro area macroeconomic aggregates. Usually, the aggregation weights which are used to produce the aggregate variable are considered as combination weights. In this work, some widely used combination weights such as equal weights and the LS estimators of the weights are analyzed. Furthermore, we apply a combination method designed for hierarchical time series and the shrinkage method to derive combination weights which use the information of the original aggregation weights. The empirical analysis yields some interesting results.

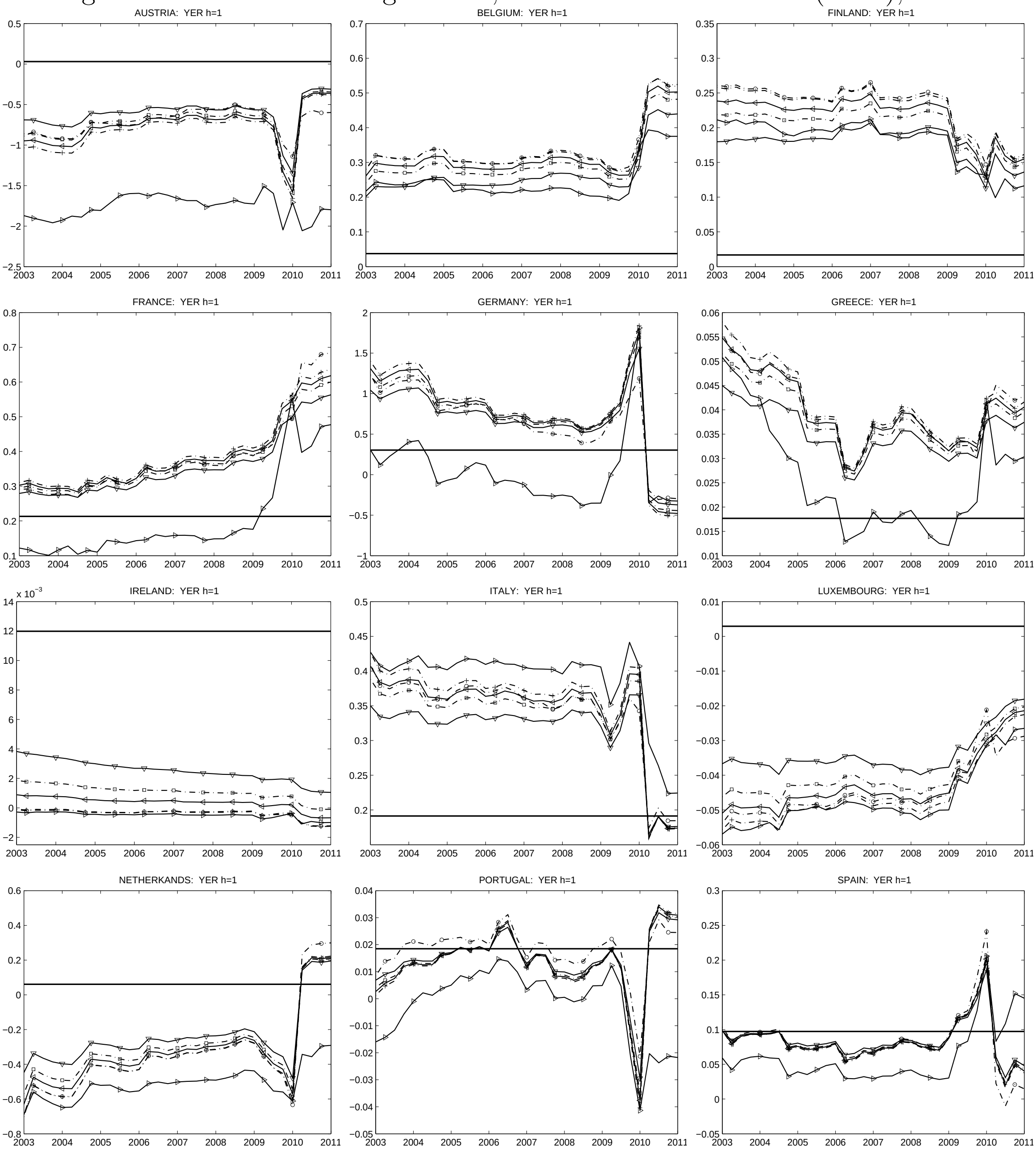
First, when forecasting the Euro area aggregate real GDP and GDP deflator, some gains can be achieved by using more flexible combination weights as for instance the LS weights and the weights suggested by the shrinkage method, because these methods are more flexible and can adapt to changes of the DGPs during the crisis period for real GDP. They are also more appropriate to take the large differences in inflation rates of member countries into account. For forecasting the aggregate CPI and long-term interest rate, using the aggregation weights to combine the country-specific forecasts beats the forecast of the aggregate variables based on their own past information at all considered horizons. Using the LS combination weights and the weights derived from the shrinkage method help to improve the forecasting accuracy of CPI and LTN in some cases.

Second, selecting only forecasts of the three largest EMU countries Germany, France and Italy for combination helps to improve the forecast precision in some cases, mainly because of their economic importance in the Euro area. In these cases, information from additional countries is not adding additional information beneficial for forecasting.

Third, if the aggregate and the country-specific data exhibit different persistence, then selecting only forecasts of those countries with highly persistent data for combination seems to improve the forecast accuracy. Future research may investigate the role of the underlying persistence in the data for the relative performance of the combination methods in more detail.

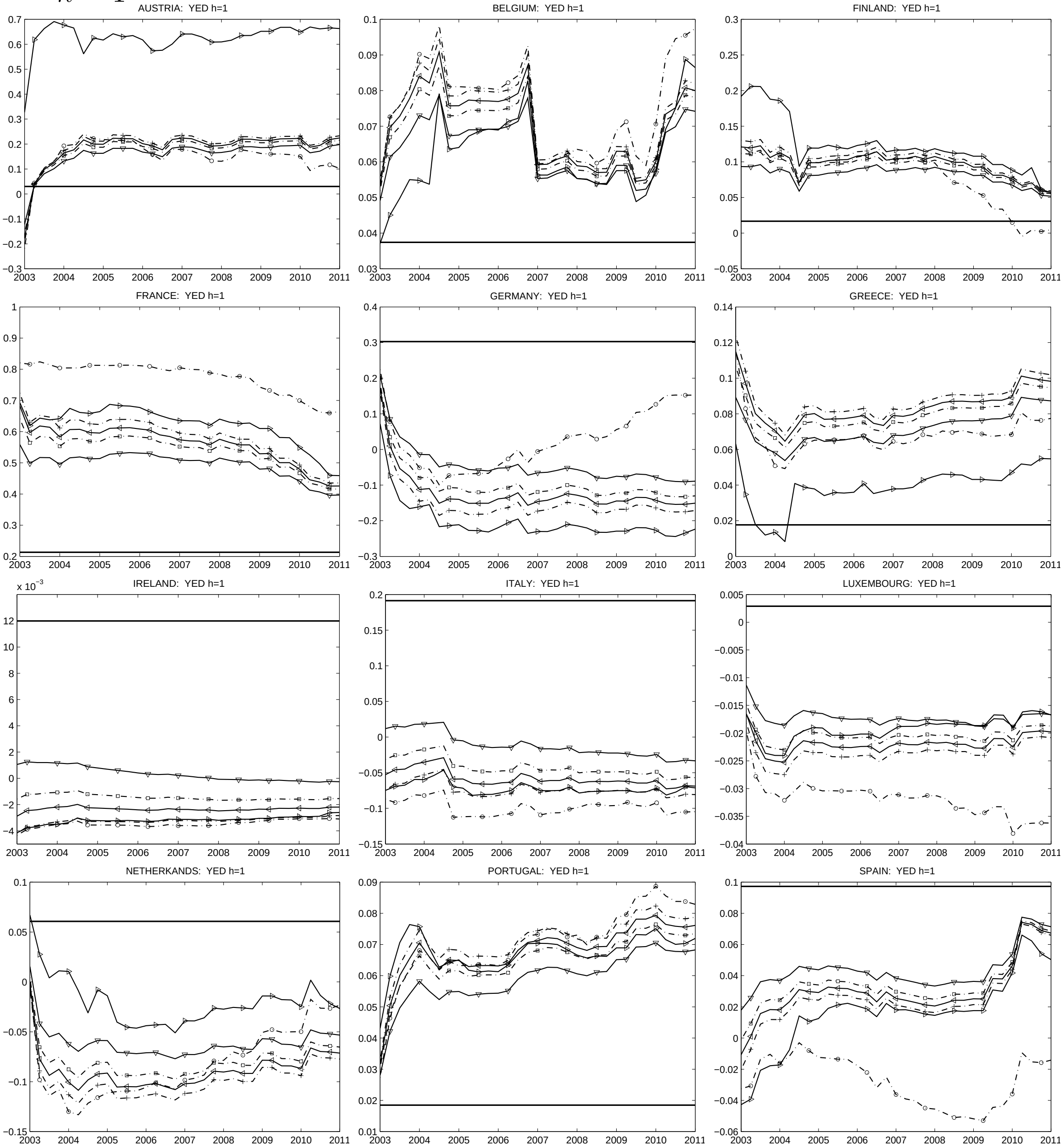
Fourth, the combination method designed for hierarchical time series does not take the actual data into account and cannot change the relative weights of the disaggregate components. Therefore, in our application, it performs quite similarly as forecasts using the original aggregation weights. Improvements regarding these issues can be considered for future research.

Figure 2: Combination weights w^{LS}, w^{S} and w^{0} for real GDP (YER), $h = 1$



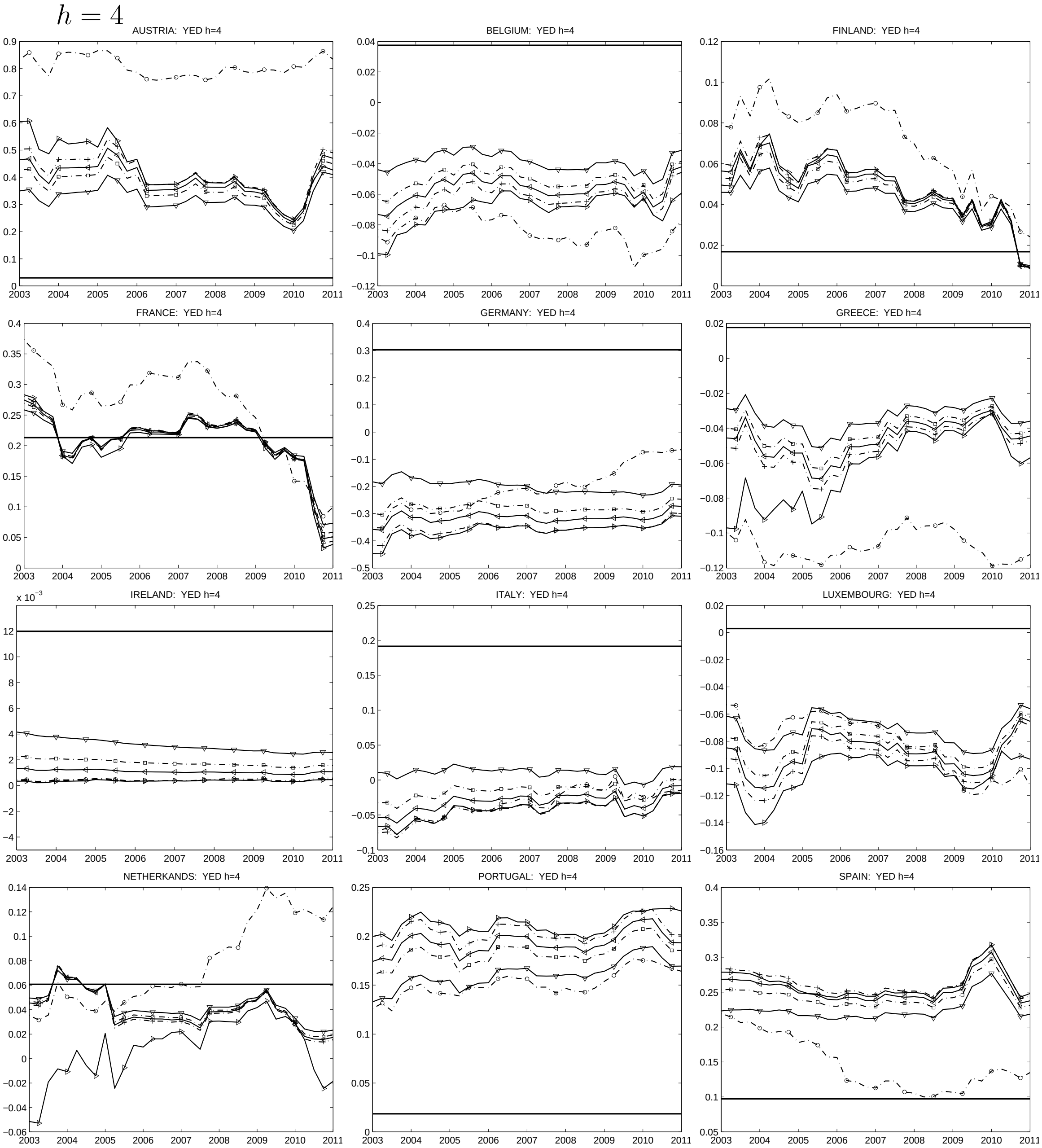
Note: The recursive out-of-sample forecasting experiment for $h = 1$ is done over the period from 2003Q1 to 2011Q1, thus we obtain 33 estimated combination weights by using the LS method and the shrinkage method. The solid horizontal line refers to the AWM weights which are fixed. The line with right-pointing triangle is for the combination weight w^{LS1}, the dotted line with plus sign for w^{LS2}, the dotted line with circle for w^{LS3}, the line with left-pointing triangle for w^{S1}, the dotted line with square for w^{S2} and the line with downward-pointing triangle for w^{S3}.

Figure 3: Combination weights w^{LS}, w^{S} and w^{0} for GDP deflator (YED), $h = 1$



Note: The recursive out-of-sample forecasting experiment for $h = 1$ is done over the period from 2003Q1 to 2011Q1, thus we obtain 33 estimated combination weights by using the LS method and the shrinkage method. The solid horizontal line refers to the AWM weights which are fixed. The line with right-pointing triangle is for the combination weight w^{LS1}, the dotted line with plus sign for w^{LS2}, the dotted line with circle for w^{LS3}, the line with left-pointing triangle for w^{S1}, the dotted line with square for w^{S2} and the line with downward-pointing triangle for w^{S3}.

Figure 4: Combination weights w^{LS}, w^{S} and w^{0} for GDP deflator (YED), $h = 4$



Note: The recursive out-of-sample forecasting experiment for $h = 4$ is done over the period from 2003Q4 to 2011Q4, thus we obtain 33 estimated combination weights by using the LS method and the shrinkage method. The solid horizontal line refers to the AWM weights which are fixed. The line with right-pointing triangle is for the combination weight w^{LS1}, the line with plus sign for w^{LS2}, the line with circle for w^{LS3}, the line with left-pointing triangle for w^{S1}, the line with square for w^{S2} and the line with downward-pointing triangle for w^{S3}.

References

Altissimo, F., L. Bilke, A. Levin, T. Mathä, and B. Mojon (2006). Sectoral and aggregate inflation dynamics in the Euro area. *Journal of the European Economic Association 4(2-3)*, 585–593.

Anderson, H. M., M. Dungey, D. R. Osborn, and F. Vahid (2011). Financial integration and the construction of historical financial data for the Euro Area. *Economic Modelling 28(4)*, 1498–1509.

Andrawis, R. R., A. F. Atiya, and H. El-Shishiny (2011). Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting 27(3)*, 870–886.

Athanasopoulos, G., R. A. Ahmed, and R. J. Hyndman (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting 25(1)*, 146–166.

Banerjee, A., M. Marcellino, and I. Masten (2005). Leading indicators for Euro-area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics 67, Supplement*, 785–813.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operational Research Quarterly 20(4)*, 451–468.

Beck, G. W., K. Hubrich, and M. Marcellino (2009). Regional inflation dynamics within and across euro area countries and a comparison with the United States. *Economic Policy 24(57)*, 142–184.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics 123(3)*, 1005–1060.

Beyer, A., J. A. Doornik, and D. F. Hendry (2001). Constructing historical Euro-zone data. *The Economic Journal 111*, F102–F121.

Bosker, E. M. (2006). On the aggregation of eurozone data. *Economics Letters 90*, 260–265.

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

Brüggemann, R., H. Lütkepohl, and M. Marcellino (2008). Forecasting Euro-area variables with German pre-EMU data. *Journal of Forecasting 27*, 465–481.

Capistran, C., C. Constandse, and M. Ramos-Francia (2010). Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling 27(3)*, 666–677.

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

Diebold, F. X. and P. Pauly (1990). The use of prior information in forecast combination. *International Journal of Forecasting 6*, 503–508.

Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model for the Euro area. Working Papers 42, European Central Bank.

Fagan, G., J. Henry, and R. Mestre (2005). An area-wide model for the Euro area. *Economic Modelling 22*, 39–59.

Fliedner, G. (1999). An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research 26(10-11)*, 1133–1149.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting 29*, 108–121.

Granger, C. W. J. (1989). Combining forecasts - twenty years later. *Journal of Forecasting 8(3)*, 167–173.

Granger, C. W. J. (1990). Aggregation of time-series variables: A survey. In T. Barker and M. H. Pesaran (Eds.), *Disaggregation in econometric modelling*, pp. 17–34. Routledge Revivals.

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

Greene, W. H. and T. G. Seaks (1991). The restricted least squares estimator: A pedagogical note. *The Review of Economics and Statistics 73(3)*, 563–567.

Grunfeld, Y. and Z. Griliches (1960). Is aggregation necessarily bad? *The Review of Economics and Statistics 42(1)*, 1–13.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal 7(1)*, 1–31.

Hendry, D. F. and K. Hubrich (2006). Forecasting economic aggregates by disaggregates. Working Papers 589, European Central Bank.

Hendry, D. F. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics 29*(2), 216–227.

Hsiao, C. and S. K. Wan (2014). Is there an optimal forecast combination? *Journal of Econometrics 178(2)*, 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.

Hyndman, R. J., A. J. Lee, and E. Wang (2014). Fast computation of reconciled forecasts for hierarchhier and grouped time series. *Working paper 17/14, Department of Econometrics and Business Statistics, Monash University*.

Küster, K. and V. Wieland (2010). Insurance policies for monetary policy in the Euro area. *Journal of the European Economic Association 8(4)*, 872–912.

Lütkepohl, H. (1984a). Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business and Economic Statistics 2(3)*, 201–214.

Lütkepohl, H. (1984b). Linear transformations of vector ARMA processes. *Journal of Econometrics 26(3)*, 283–293.

Lütkepohl, H. (2010). Forecasting aggregated time series variables: A survey. *OECD Journal: Journal of Business Cycle Measurement and Analysis 2010(2)*, 37–62.

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

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.

Schwarzkopf, A. B., R. J. Tersine, and M. J. Morris (1988). Top-down versus bottom-up forecasting strategies. *International Journal of Production Research 26(11)*, 1833–1843.

Shlifer, E. and R. W. Wolff (1979). Aggregation and proration in forecasting. *Management Science 25(6)*, 594–603.

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. H. and M. W. Watson (2002a). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.

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

Swanson, N. R. and T. Zeng (2001). Choosing among competing econometric forecasts: Regression-based forecast combination using model selection. *Journal of Forecasting 20*, 425–440.

Tiao, G. C. and I. Guttman (1980). Forecasting contemporal aggregates of multiple time series. *Journal of Econometrics 12(2)*, 219–230.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, Chapter 4, pp. 135–196. North Holland, Amsterdam.

Wallis, K. F. (2011). Combining forecasts - forty years later. *Applied Financial Economics Volume 21, Issue 1-2, 2011 21(1-2)*, 33–41.

Wei, W. W. S. and B. Abraham (1981). Forecasting contemporal time series aggregates. *Communications in Statistics - Theory and Methods 10(13)*, 1335–1344.

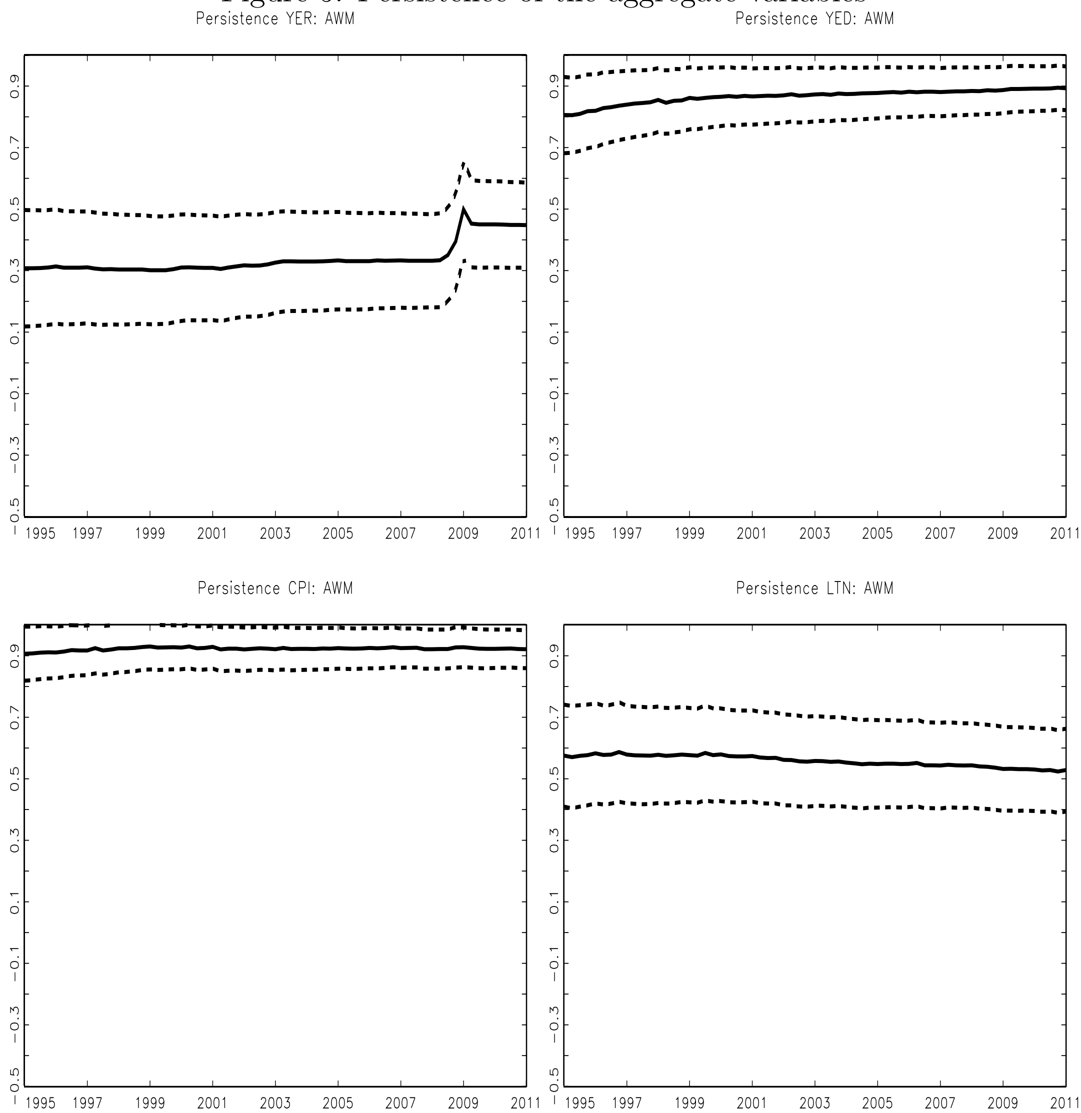
Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics 146*, 329–341.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In P. K. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*. North Holland, Amsterdam.

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

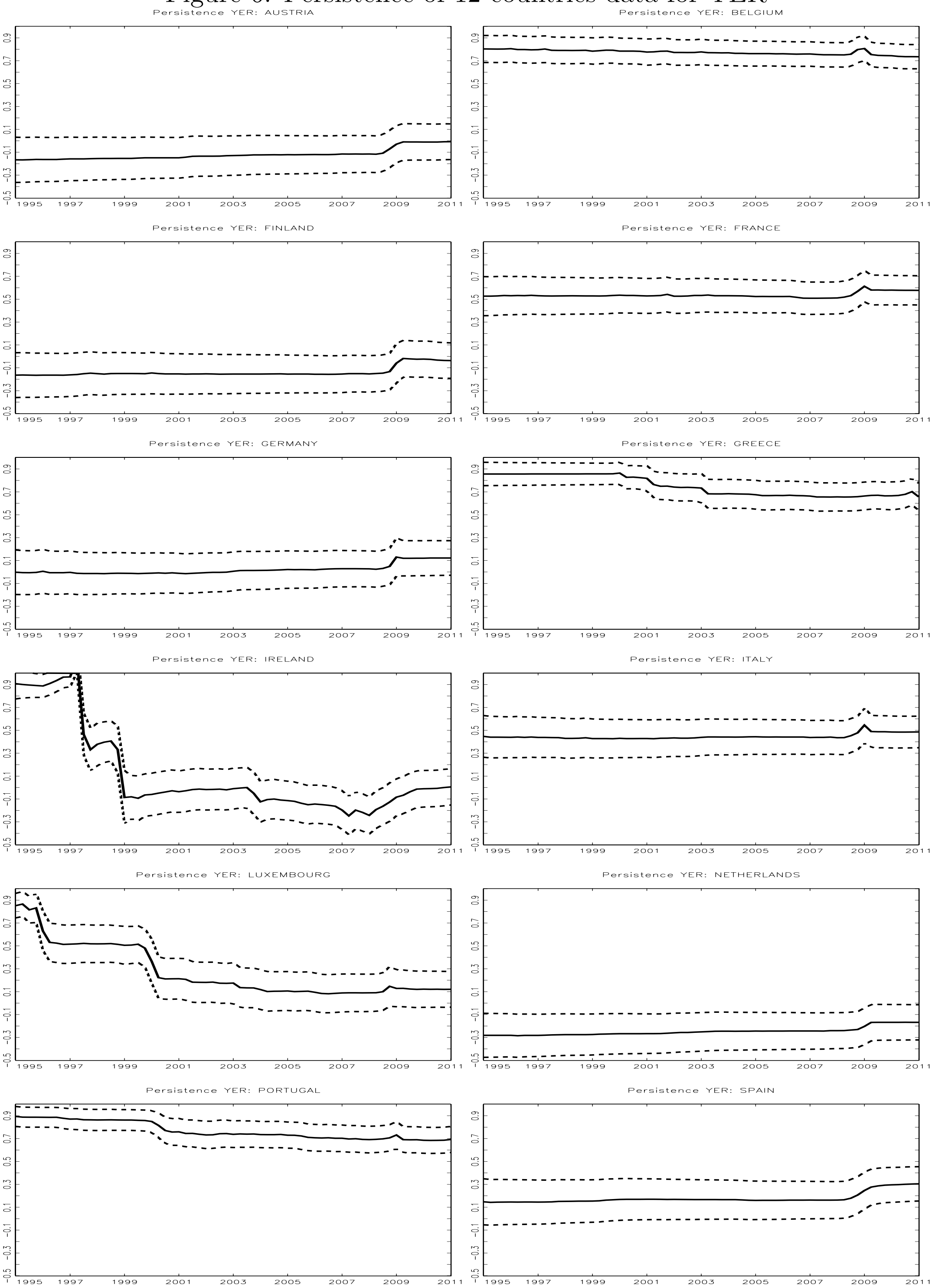
A Plots of the persistence of the data

Figure 5: Persistence of the aggregate variables



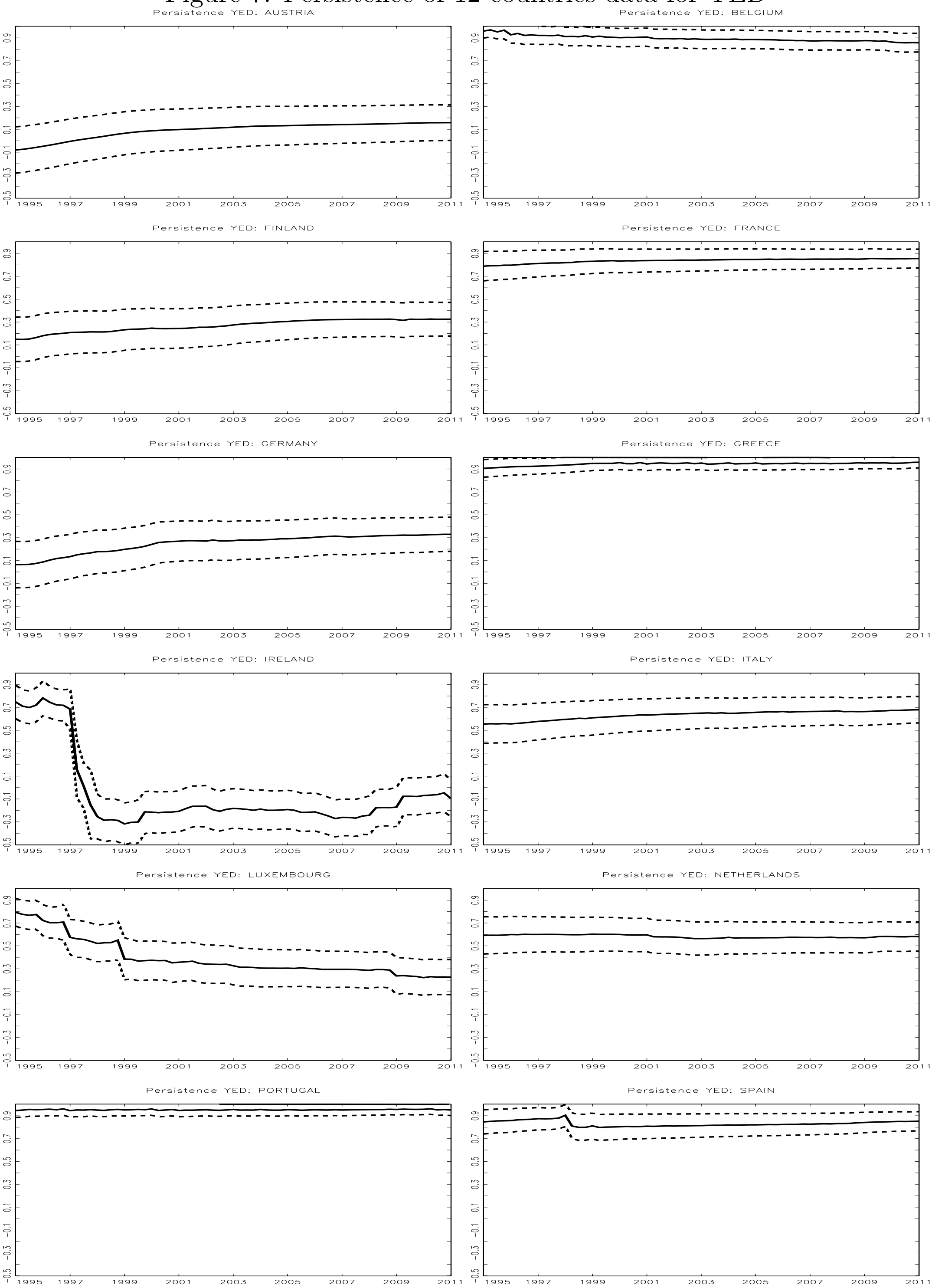
Note: Recursive estimated AR(1) coefficients with 95% confidence intervals.

Figure 6: Persistence of 12 countries data for YER



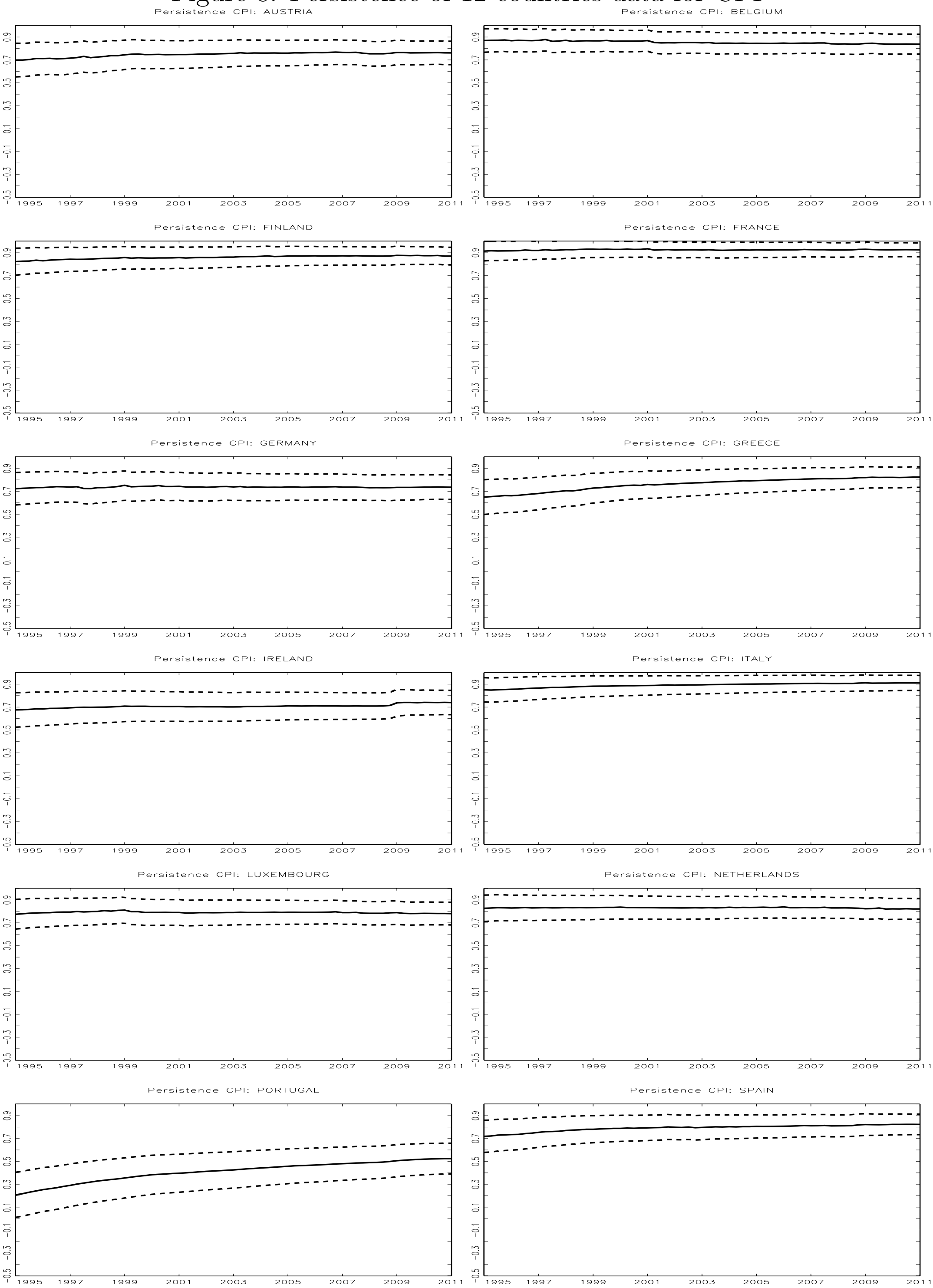
Note: Recursive estimated AR(1) coefficients with 95% confidence intervals.

Figure 7: Persistence of 12 countries data for YED



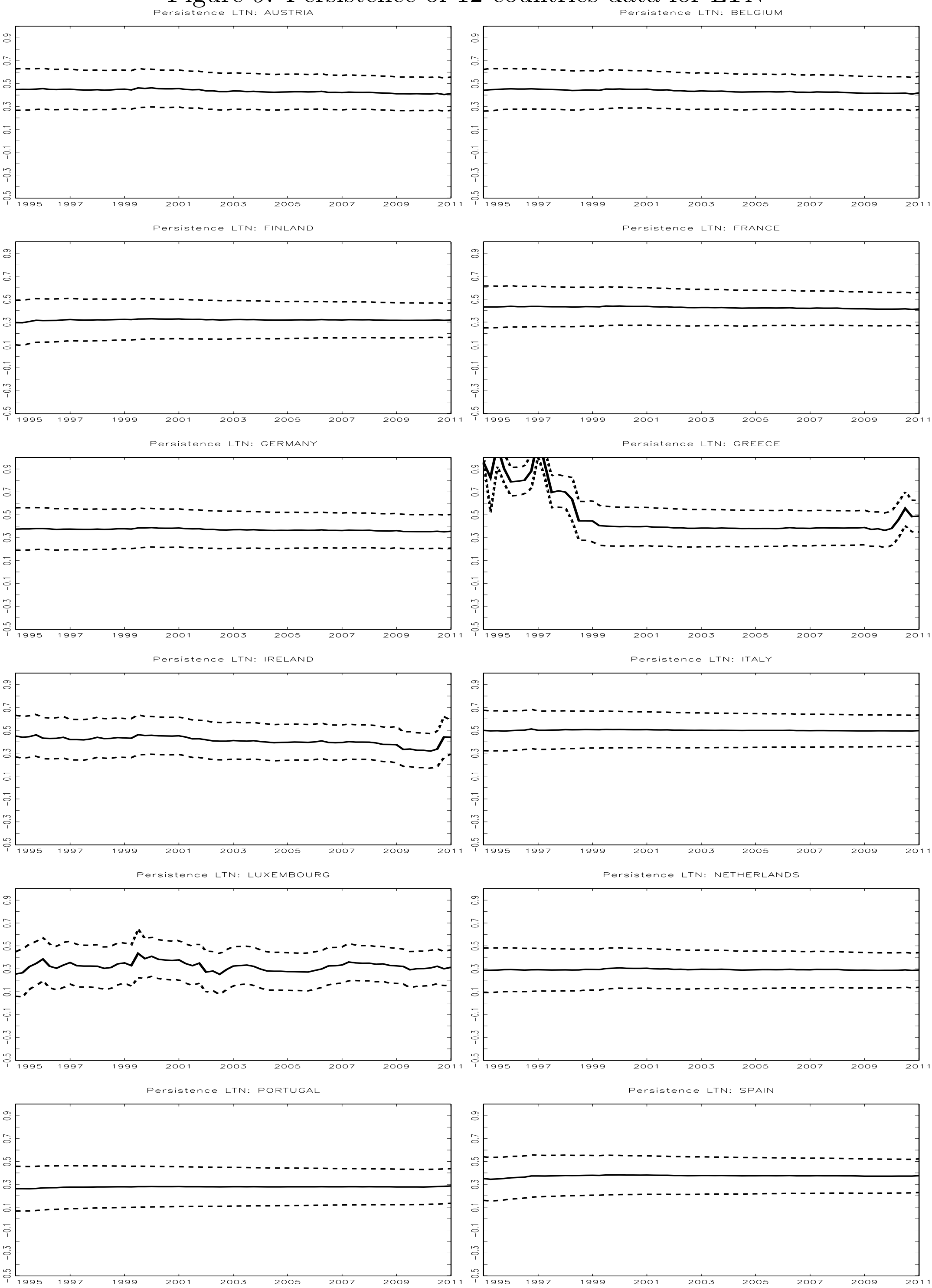
Note: Recursive estimated AR(1) coefficients with 95% confidence intervals.

Figure 8: Persistence of 12 countries data for CPI



Note: Recursive estimated AR(1) coefficientswith 95% confidence intervals.

Figure 9: Persistence of 12 countries data for LTN



Note: Recursive estimated AR(1) coefficients with 95% confidence intervals.

UNIVERSITY OF KONSTANZ
Department of Economics

Universitätsstraße 10
78464 Konstanz
Germany

Phone: +49 (0) 7531-88-0
Fax: +49 (0) 7531-88-3688

www.wiwi.uni-konstanz.de/econdoc/working-paper-series/