Forecasting Euro-area macroeconomic variables using a factor model approach for backdating

Brüggemann, Ralf; Zeng, Jing

CENTER FOR QUANTITATIVE METHODS AND SURVEY RESEARCH

CMS Discussion Paper 2010-04

Forecasting Euro-Area Macroeconomic Variables Using a Factor Model Approach for Backdating

Ralf Brüggemann*
Jing Zeng*

*University Konstanz, Department of Economics

http://www.uni-konstanz.de/cms

Forecasting Euro-Area Macroeconomic Variables Using a Factor Model Approach for Backdating*

Ralf Brüggemann† Jing Zeng‡

February 15, 2010

Abstract

We suggest to use a factor model based backdating procedure to construct historical Euro-area macroeconomic time series data for the pre-Euro period. We argue that this is a useful alternative to standard contemporaneous aggregation methods. The paper investigates for a number of Euro-area variables whether forecasts based on the factor-backdated data are more precise than those obtained with standard area-wide data. A recursive pseudo-out-of-sample forecasting experiment using quarterly data and a forecasting period 2000Q1-2007Q4 is conducted. Our results suggests that some key variables (e.g. real GDP and inflation) can indeed be forecasted more precisely with the factor-backdated data.

Keywords: forecasting, factor model, backdating, European monetary union, constructing EMU data

JEL classification: C22, C53, C43, C82

*We thank participants in the DIW Macroeconometric Workshop, December 17-18, 2009 for helpful comments.

†Universität Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: ralf.brueggemann@uni-konstanz.de

‡Universität Konstanz, Department of Economics, Box 129, 78457 Konstanz, Germany, email: jing.zeng@uni-konstanz.de

1 Introduction

With the creation of European Monetary Union (EMU), the focus of macroeconomic analysis has shifted towards the analysis of the Euro area as a whole. Econometric models for area-wide variables have been used for forecasting and structural analysis. As actual EMU time series data are only available from 1999 onwards, synthetic time series data for the pre-EMU period are in use. Often the construction of historical (pre-EMU) Euro area data is based on contemporaneous aggregation of time series from the EMU member countries. Different aggregation methods have been suggested in the literature and Marcellino (2004) points out a number of drawbacks inherent in these methods. The choice of a particular aggregation method is a very important practical issue that impacts any following econometric analysis. For instance, Bosker (2006) illustrates that estimated cointegration parameters change substantially with the choice of the aggregation method. Given these drawbacks of standard methods, it is worth to consider the merits of alternatives to aggregation. In this paper, we therefore consider a factor model based alternative to the standard method of contemporaneous aggregation and analyze the usefulness of this approach in forecasting Euro area aggregates.

One of the standard aggregation methods suggested in the literature has been discussed by Fagan, Henry & Mestre (2001, 2005). Their approach has been used to create a database of historical euro-area time series data for estimating the Area Wide Model (AWM) in use at the European Central Bank (ECB).[1] The AWM data is based on cross-country aggregation of log-level variables with fixed weights (referred to as FHM weights). The FHM weights are obtained as shares of GDP at constant 1995 prices. Anderson, Dungey, Osborn & Vahid (2007) point out that the use of fixed weights will tend to undervalue the importance of the countries, which hold a leading role in the European markets and suggest extending the FHM weights with a sliding factor which measures the relative distance from economic integration to EMU. Using fixed weights may also be problematic because it does not

[1] Updates of this database is available from the Euro Area Business Cycle Network (EABCN) at http://www.eabcn.org/

take changes in exchange rates between member countries into consideration. Therefore, Beyer, Doornik & Hendry (2001) suggest to aggregate growth rates of the variables with time-varying weights based on previous period's real GDP share (henceforth BDH weights) and find that in their method the aggregates of the individual deflators correspond to the deflator of the aggregate. Recently, Beyer & Juselius (2009) show that results based on BDH weights are sensitive to the choice of base year and therefore suggest to use weights based on previous period's nominal GDP. None of the proposed methods seems optimal in all respects.

Alternatives to standard aggregation have also been considered in the literature. For instance, Brüggemann & Lütkepohl (2006) and Brüggemann, Lütkepohl & Marcellino (2008) argue that the use of synthetically constructed, aggregated data is inappropriate especially in the presence of structural changes induced by adjustment processes required in some countries prior to EMU in order to satisfy the Maastricht criteria. They suggest a representative country approach which combines German data until 1998 with actual Euro-area data after 1999. They find that at least for some variables like interest rates and prices using German data rather than aggregated EMU data for the pre-EMU period is preferable when forecasts of EMU aggregates are of interest.

This paper proposes to use another alternative method for constructing historical Euro-area data. We extend the idea put forward in Angelini & Marcellino (2007), where a factor based approach is used to construct time series of macroeconomic variables for unified Germany prior to 1991. In the factor model approach, a small number of factors are extracted from a large set of time series from individual EMU member countries using the Stock & Watson (2002a) principal component based estimators. The estimated relation between the factor time series and the actual Euro-area time series of interested is used to construct time series data for the pre-EMU period. This method is referred to as factor-backdating. Advantages of this method include its ability to use more time series information than standard aggregation methods and its ability to handle situations with missing time series data in some of the cross-sectional units (countries). Against the background of future EMU enlargement and the doubtful quality of historical data in some of the future member countries, the factor-backdating procedure may be an attractive and useful alternative to standard aggregation methods.

We analyze the usefulness of this approach in forecasting a number of macroeconomic Euro-area variables by conducting a forecast comparison. We compare the accuracy of forecasts based on models that use different historical Euro-area time series. In particular, we compare forecasts based on pre-EMU data from the AWM database in use at the ECB to forecasts based on data obtained from the factor-backdating procedure. Our paper is related to work by Marcellino (2004), which also includes a forecasting comparison for EMU macroeconomic time series based aggregated data constructed by Fagan, Henry & Mestre (2005). However, the focus in Marcellino (2004) is on the forecasting performance of different forecasting methods, not on different data. Our work is also related to the study by Brüggemann et al. (2008) who investigate whether German data before the Euro period contain the same or more information for forecasting than the aggregated data by comparing linear and nonlinear forecasting methods. They find that at least for nominal and monetary variables German data results in superior forecasts.

Our study uses a number of linear and nonlinear forecasting methods and models. In particular, we include variants of linear autoregressive models as well as non-linear smooth transition regression models. These forecasting models have also been used in e.g. Stock & Watson (1999), Marcellino (2004) and Brüggemann et al. (2008). Variables included in our comparison are real GDP, the GDP deflator, a consumer price index, short- and long-term interest rates as well as the exchange rate.

The structure of the remaining paper is as follows. In Section 2, the factor-backdating approach is presented. The forecasting methods are discussed in Section 3, before the data are described in Section 4. Section 5 discusses the results from our forecasting comparison and Section 6 concludes.

2 Factor-based backdating

As an alternative to standard contemporaneous aggregation methods, we suggest to use a factor-based approach to backdate historical data for the Euro-area. In this factor-backdating procedure, a small number of common factors is extracted from a possibly large set of time series data coming from individual Euro-area countries. Using the period where both, the information on the actual aggregated Euro-area time series (for the period after 1999) and the extracted factor time series are available, we estimate the relation between the unobserved factors and the area-wide aggregate. This information is in turn used to backdate historical Euro-area data. A detailed description of the factor backdating procedure is given in the following.[2]

Starting point is a factor model representation discussed by Stock & Watson (2002a, 2002b). In their approach the N-dimensional stationary time series X_t is driven by a small number of K unobserved common factors F_t and an idiosyncratic component e_t, i.e. the vector of time series may be written as

$$X_t = \Lambda F_t + e_t, \quad t = 1, \ldots, T, \tag{2.1}$$

where X_t is a $N \times 1$ vector, Λ is a $N \times K$ matrix of factor loadings, F_t is the $K \times 1$ vector of common factors and e_t is an $N \times 1$ vector of idiosyncratic components. Prior to the backdating procedure, the common factors have to be extracted from the time series data. Estimation of the factors is done using a classical static principle components on $\tilde{X}_t$, which is obtained by standardizing X_t to have mean zero and unit variance. This procedure gives a K-dimensional time series of common factors, denoted as $\{\hat{F}_t\}_{t=1}^T$. It can be shown that under mild conditions the principal components of $\tilde{X}_t$ are consistent estimators of the true unobservable factors (see e.g. Stock & Watson (2002a) for details). In applications, the variables in X_t are transformed to stationarity if necessary. The choice of the number of factors may be based on suitable criteria (see e.g. Bai & Ng (2002)).

In our application, the vector X_t consists of a number of time series coming from the individual Euro-area member countries and we extract the factor time series using data over the entire sample period. In the second step, we relate the factor time series to the area-wide macroeconomic series of interest using a period, where observations on both are available. To be more precise, we regress the (stationarity transform) of the Euro-area-wide

[2]This approach has been used in the context of backdating German data by Angelini & Marcellino (2007).

series of interest, denoted as y_t^{EMU}, on the estimated factor time series $\hat{F}_t$ over the period from 1999Q1 to 2007Q4, i.e. we use

$$y_t^{EMU} = \beta_0 + \beta_1 \hat{F}_{1t} + \ldots + \beta_K \hat{F}_{Kt} + \varepsilon_t, \quad t = 1999Q1, \ldots, 2007Q4. \tag{2.2}$$

and estimate the parameters $\beta_0, \ldots, \beta_K$ by OLS. In the third step of our procedure, we use the estimated parameters $\hat{\beta}_0, \ldots, \hat{\beta}_K$ to backdate the (historical) area-wide time series for the periods before 1999 by:

$$\hat{y}_t^{EMU} = \hat{\beta}_0 + \hat{\beta}_1 \hat{F}_{1t} + \ldots + \hat{\beta}_K \hat{F}_{Kt}, \quad t = 1970Q1, \ldots, 1998Q4. \tag{2.3}$$

This factor model approach has several advantages: While it uses the information from all member countries like standard aggregation methods do, it avoids the difficulty to choose the appropriate aggregation weights. In the factor-backdating method, the weights are obtained in a data driven way. Moreover, standard aggregation methods typically only use the country information on the one variable that is aggregated. In the factor based approach, X_t can in principle include many other variables as well. For instance, when constructing Euro-area data for the overall consumer price index, the vector X_t used in the factor backdating may include the time series of consumer price subindexes or from other price indices as well. Therefore, the information content of an area-wide time series obtained by the described factor method may be greater than in a time series obtained by standard aggregation methods. The factor-based method is also suitable to handle missing time series observations in some cross-sectional units, a situation that occurs often when constructing Euro-area aggregates. Even if there are some missing observations in some of the cross-sections, the common factors can still be extracted by using the expectation-maximization (EM) algorithm (see e.g. the discussion in Appendix A of Stock & Watson (2002a) and in Angelini, Henry & Marcellino (2006)). Against the background of future EMU enlargement and the doubtful quality of historical data in some of the future member countries, the factor backdating procedure may be an attractive and useful alternative to standard aggregation methods.

The usefulness of the suggested approach in forecasting is investigated in the remaining part of the paper.

3 Forecasting Methods and Evaluation

3.1 Forecasting Methods

The forecasting methods used in this work are similar to those discussed by Stock & Watson (1999), Marcellino (2004) and Brüggemann et al. (2008). Thus, only a brief description of the different methods are given in the following.

In our forecasting exercise we are interested in forecasting the EMU aggregate of some variable of interest h periods ahead. We denote this variable as y_t^{EMU}. Depending on the integration properties of this variable, the forecasting model is either specified for the level y_t^{EMU} or for the first difference $\Delta y_t^{EMU} = y_t^{EMU} - y_{t-1}^{EMU}$. To make the forecast errors comparable across both cases, we specify forecasting models for the variable y_{t+h}^h, where $y_{t+h}^h = y_{t+h}^{EMU}$, when the variable is stationary and $y_{t+h}^h = y_{t+h}^{EMU} - y_t^{EMU}$, when the variable is integrated of order one ($I(1)$). h denotes the forecasting horizon. All considered forecasting methods can be written as

$$y_{t+h}^h = f(Z_t; \theta_{ht}) + \varepsilon_{t+h}, \tag{3.1}$$

where Z_t is the vector of explanatory variables, θ_{ht} is a vector of possibly time-varying parameters and ε_t is an error term. The h-step ahead forecast is given by replacing the unknown parameter vector θ_{ht} by an estimate and hence,

$$\hat{y}_{t+h}^h = f(Z_t; \hat{\theta}_{ht}), \tag{3.2}$$

and the h-step forecast error is

$$e_{t+h} = y_{t+h}^h - \hat{y}_{t+h}^h = y_{t+h} - \hat{y}_{t+h}. \tag{3.3}$$

We use $h = 1$, $h = 2$ and $h = 4$ as forecasting horizons. In the case of multi-step predictions, we use the so-called 'h-step ahead projection', which is also known as the 'direct forecast' approach (see e.g. Clements & Hendry (1996)). In other words, different forecasting models are fitted for different forecasting horizons. In comparison to the 'iterated multi-step forecasts', the direct forecasting method is advantageous in the context of nonlinear models as simulation from nonlinear models is avoided.

We use linear autoregressive models as well as non-linear smooth transition regression models. The model variants are briefly described in the following:

Autoregressions (AR). This simple linear forecasting method has the form

$$y^h_{t+h} = \mu_t + \beta' Z_t + \varepsilon_{t+h}. \tag{3.4}$$

If y^{EMU}_t is treated as a stationary variable, then $Z_t = (y_t, \dots, y_{t-p+1})'$, otherwise $Z_t = (\Delta y_t, \dots, \Delta y_{t-p+1})'$, where the superscript EMU has been dropped to simplify the notation. In the deterministic component μ_t a constant or a constant and a time trend can be included. Three variants of lag lengths are considered: a fixed number of lags $p = 4$; lag length selected by AIC $(0 \leq p \leq 4)$; lag length selected by BIC $(0 \leq p \leq 4)$. Since the variable y^{EMU}_t can be treated as stationary, or as $I(1)$, or a unit root pre-test may be used, there are 18 model variants in this class. The different variants are denoted as A1-A18 and listed in Panel A of Table 1.

Logistic smooth transition autoregression (LSTAR). This nonlinear forecasting method is of the form

$$y^h_{t+h} = \alpha' Z_t + d_t \beta' Z_t + \varepsilon_{t+h}. \tag{3.5}$$

As in the autoregressive models, $Z_t = (1, y_t, \dots, y_{t-p+1})'$ if y_t is treated stationary or $Z_t = (1, \Delta y_t, \dots, \Delta y_{t-p+1})'$ if y_t is integrated. The term d_t is a logistic function $d_t = 1/[1+\exp(\gamma_0+\gamma_1\zeta_t)]$. The value of the so-called smoothing parameters γ_1 determines the shape of parameter change over time. For $\gamma_1 = 0$, the model becomes linear, while for large values of γ_1 the model tends to a self-exciting threshold model, see e.g. Granger & Teräsvirta (1993) and Teräsvirta (1998) for details. ζ_t is the transition variable and in the considered variants may depend on current and past y_t. For models specified in levels, the following five alternatives are used for ζ_t: $\zeta_t = y_t$; $\zeta_t = y_{t-1}$; $\zeta_t = y_{t-3}$; $\zeta_t = y_t - y_{t-2}$; $\zeta_t = y_t - y_{t-4}$. The choice of the transition variable follows Marcellino (2004). For models specified in first differences, the following five alternatives are used for ζ_t: $\zeta_t = \Delta y_t$; $\zeta_t = \Delta y_{t-1}$; $\zeta_t = \Delta y_{t-3}$; $\zeta_t = y_t - y_{t-2}$; $\zeta_t = y_t - y_{t-4}$. The lag length p of the model is fixed 2 in some models, while in some other variants we use AIC and BIC to select from a choice of models with $p = 1, 2, 4$ and all possible ζ_t mentioned above. This gives a total of 12 different LSTAR model variants, denoted as L1-L12. The models used are listed in Panel B of Table 1.

3.2 Forecasting Comparison

In this work we do not focus on the comparison of forecasting methods, but investigate whether the AWM data or factor-backdated data is preferable for making forecasts. For this purpose, we conduct a recursive pseudo-out-of-sample forecasting experiment and look at forecasting precision at horizons $h = 1, 2$ and 4. In our experiment, the initial estimation period covers 1970Q1-1999Q4, i.e. $T = 120$ observations. The forecast period is $2000Q1 - 2007Q4$ and consists of 32 quarters. To mimic the behavior of a forecaster, the unit root pre-tests, model selection and estimation are repeated once a new observation is added to the estimation period.

To compare the forecasting performance, the mean squared forecast error (MSFE) is used as loss function. For forecast horizon h, model m and variable n with type of data j it can be defined as:

$$MSFE^h_{n,m,j} = \frac{1}{33 - h} \sum_{t=T+h}^{T+33-h} (e_{t,n,m,j})^2, \tag{3.6}$$

where the forecast error is $e_{t+h} = y^h_{t+h} - \hat{y}^h_{t+h}$. To simplify the comparison, each MSFE obtained from the factor-backdating approach, denoted as, $MSFE^h_{n,m,F}$ will be expressed relative to the MSFE obtained from models based on AWM data, denoted as $MSFE^h_{n,m,AWM}$. Thus, if the relative MSFE is less than one, the forecasts based on factor backdated data are more precise than forecasts based on AWM data.

4 Data

Our forecasting comparison includes six Euro-area macroeconomic variables on a quarterly frequency: real GDP (YER), the GDP deflator (YED), the consumer price index (CPI), the exchange rate against the US-Dollar (EER) and short- and long-term interest rates (STN and LTN). The mnemonics correspond to those in the AWM database.

The first set of area-wide time series corresponds to data obtained from the AWM database maintained at the Euro Area Business Cycle Network[3]. As mentioned before, the AWM data is based on cross-country aggregation of log-level variables with fixed weights. The aggregation methods is the one used in Fagan, Henry & Mestre (2001) and Fagan et al. (2005). This AWM data is now in widespread use, e.g. within the ECB for estimating econometric models.[4] Quarterly data for a period from 1970Q1 to 2007Q4 is used in the following and the corresponding time series plots are given in Figure 1.

As an alternative we consider a set of time series obtained from the backdating procedure described in Section 2. For this procedure the individual member countries' time series data are taken from the OECD quarterly national accounts database and are available for a period from $1970Q1$ to $2007Q4$.[5] Figure 1 depicts for each considered variable time series of the three largest Euro-area member states Germany, France, and Italy. For some variables, like e.g. price measures and interest rates, the time series plots reflect quite different developments in the three countries. In the forecasting exercise, we consider the log-transform of real GDP, the GDP deflator, the consumer price index and the exchange rate, while short- and long-term interest rates are not transformed. Time series on the variables for all twelve considered countries are characterized by trends and their is evidence that the series can be characterized as $I(1)$ processes.[6] Therefore, the first difference of the variables enters the vector X_t, which after standardization is used to estimate the common factors. In this study, the factors are extracted from a set of country time series data that consists of variables corresponding to the aggregate of interest. For instance, when backdating area-wide real GDP growth, the factors are extracted from a set X_t that only includes real

[3]http://www.eabcn.org

[4]It should be noted, however, that the AWM database is not an official ECB database.

[5]Twelve eurozone countries are considered: Austria, Belgium, Finland, France, Germany, Greece, Italy, Ireland, Luxembourg, Netherlands, Portugal, Spain. The data are obtained via Thomson Datastream.

[6]A formal unit root analysis has been conducted and the results are available on request.

GDP growth from the member countries and no other variables are taken into account.

For each variable, we determine the number of factors by looking at the percentage of variance explained by each principal component. We use a type of 'elbow criteria' and use the minimum number of factors that explains more than 70% of the variance. Table 2 reports for each variable the number of factors used in the backdating procedure. Not surprisingly, the number of factors needed varies with the considered variables. For instance, for backdating real GDP growth (YER) six factors are needed, which might reflect the fairly heterogeneous developments in the real economies of the Euro-area member states. In contrast, only one factor is selected for the exchange rate (EER), which may be due to the German dominance within the European Monetary System (EMS). The factor backdating procedure described in Section 2 is applied for the six mentioned variables, where factor extraction is over the period 1970Q1-2007Q4 and the corresponding backdating is for the period 1970Q1-1998Q4. As we treat all variables as $I(1)$, the backdating is done on the first differences of the respective variables. From the backdated changes (and growth rates) we compute the respective level of the time series. This approach gives a set of six factor-backdated Euro-area time series and plots of these series are given in Figure 2.

Comparing the area-wide series from the AWM database with those obtained by the factor backdating procedure (see Figure 2) shows that for the pre-Euro period both methods lead to time series that have similar trending behavior. Nevertheless, the medium and short-term fluctuations are typically quite different, which in turn may have an impact on the forecasting performance. From 1999Q1 onwards, time series from both approaches are identical because both use actual Euro-area data.

5 Results

The results from our forecasting comparison for all six variables are presented in Figures 3 to 8. For each variable and forecasting model variant (see Table 1), we report the MSFE of the model based on factor-backdated data relative to the MSFE of the corresponding model based on AWM data. Results for forecasting horizons $h = 1, 2$ and 4 are reported in the upper, middle and bottom panel of the corresponding figure, respectively.

Figure 3 shows the results for real GDP (YER). For $h = 1$, the overall forecasting performance of models based on factor-backdated data is comparable to models based on AWM models. Nevertheless, in some of the forecasting models the use of factor-backdated is beneficial with some sizable gains in forecasting precision. For longer forecasting horizons, the results are more clear-cut. For $h = 2$ and $h = 4$, using factor-backdated data is beneficial in most of the considered model specifications with substantial reductions in MSFE in some of the forecasting models.

Figure 4 shows the results for the GDP deflator (YED). For $h = 1$, we find that factor-backdated data leads to more precise forecasts in all considered linear autoregressive model variants, while for most nonlinear models using aggregated AWM data seems to beneficial. At higher forecasting horizons we also find that most linear models based on factor-based data outperform the corresponding forecasting models based on AWM data but the gains in MSFEs tend to be smaller (especially at $h = 4$).

The results for the consumer price index (CPI, see Figure 5) is rather mixed. On the one hand, using the factor-backdating approach in linear models typically leads to comparable (or lower) forecasting accuracy as using AWM data, although some factor-backdated models can be identified that have relative MSFEs below one. On the other hand, some nonlinear LSTAR models perform clearly better when the factor-backdated data is employed (see e.g. the results for $h = 1$ and $h = 4$).

For the exchange rate variable (EER, see Figure 6) using the factor-backdated data is generally not beneficial when the focus is on linear models. Most of the relative MSFEs are around one, indicating that both data variants perform equally well in predicting the exchange rate. Interestingly, some of the LSTAR variants with backdated time series outperform their AWM counterparts.

For the considered long- and short term interest rates (LTN) and (STN) a similar picture emerges (see Figures 7 and 8). Although some linear models based on backdated data perform slightly better than their AWM data counterparts, we note that potential gains in linear model variants are typically negligible. Moreover, for the short-term rate we identify some AWM-data models that outperform corresponding models that use backdated time series. Nevertheless, for the long-term rate gains from using backdated data are typically visible when LSTAR models are used.

Table 3 shows for each of the considered variables and for each forecasting horizon the three best performing model/data variants together with the corresponding relative MSFEs.[7] This is an alternative way to summarize the results of our forecasting comparison and can be employed to judge the usefulness of the factor-backdating approach. The results in Table 3 indicate, for instance, that at all horizons, the two best forecasting models for real GDP (YER) are linear AR models that use factor-backdated pre-euro data. A similar picture arises for the GDP deflator. Thus, using factor-based backdated time series is beneficial for forecasting GDP growth, the aggregate price level and inflation, respectively. In addition, for the other considered variables, a model that used backdated time series is among the top three performing models and is often the overall best model.

Overall, our results indicate that for some key variables like real GDP and inflation using factor-backdated data for the pre-euro period is a useful strategy when forecasts are of interest.

6 Conclusion

In this paper we have suggested to use a factor model based backdating procedure to construct historical Euro-area macroeconomic time series data for the pre-Euro period. We argue that this is a useful alternative to standard contemporaneous aggregation methods as it may be used in situations where time series data from some cross-sectional units is missing or not available in the desired quality. Against the background of future EMU enlargement and the doubtful quality of historical data in some of the future member countries, the factor-backdating procedure may be an attractive and useful alternative to standard aggregation methods.

[7]Note that we have now used the same benchmark model for all models in order facilitate the comparison across different specifications. The benchmark is an AR(4) with constant, specified in levels for the variable using AWM data.

We have conducted a recursive pseudo-out-of-sample forecasting experiment to investigate for a number of Euro-area variables whether forecasts based on the factor-backdated data are more precise than those obtained with standard area-wide (AWM) data. A forecasting period 2000Q1-2007Q4 has been used. Our results suggests that some key variables (e.g. real GDP and inflation) can indeed be forecasted more precisely with the factor backdated data.

Overall, our results indicate that for some important variables the factor-backdating procedure is a valuable method to construct time series data for the Euro-area.

References

Anderson, H., Dungey, M., Osborn, D. R. & Vahid, F. (2007). Constructing historical euro area data, *CAMA Working Papers, Australian National University, Centre for Applied Macroeconomic Analysis* .

Angelini, E., Henry, J. & Marcellino, M. (2006). Interpolation and backdating with a large information set, *Journal of Economic Dynamics and Control* **30**: 2693–2724.

Angelini, E. & Marcellino, M. (2007). Econometric analyses with backdated data - unified Germany and the euro area, *ECB Working Paper Series, No. 752* .

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **71**(1): 135–172.

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

Beyer, A. & Juselius, K. (2009). Does it matter how to measure aggregates? the case of monetary transmission mechanisms in the euro area, *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Castle, J. L. and Shepard, N. (eds)*, Oxford University Press.

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

Brüggemann, R., Lütkepohl, H. & Marcellino, M. (2008). Forecasting euro-area variables with german pre-EMU data, *Journal of Forecasting* **27**: 465–481.

Brüggemann, R. & Lütkepohl, H. (2006). A small monetary system for the euro area based on German data, *Journal of Applied Econometrics* **21**(6): 683–702.

Clements, P. M. & Hendry, D. F. (1996). Multi-step estimation for forecasting, *Oxford Bulletin of Economics and Statistics* **58**: 657–684.

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

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

Granger, C. W. J. & Teräsvirta, T. (1993). Modelling non-linear economic relationships, *Oxford: Oxford University Press* .

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

Stock, J. H. & Watson, M. W. (1999). A comparison of linear and non-linear univariate models for forecasting macroeconomic time series, *Cointegration, causality, and forecasting: A festschrift in honor of Clive W. J. Granger. Engle, R. and White, R. (eds)*, Oxford University Press.

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

Stock, J. H. & Watson, M. W. (2002b). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* **97**: 1167–1179.

Teräsvirta, T. (1998). Modelling economic relationships with smooth transition regressions, *Handbook of Applied Economic Statistics. Ullah, A. and Giles, D. E. A. (eds)*, New York: Marcel Dekker.

Table 1: Forecasting models

A. Linear models: Autoregressive models (18 variants)	
A1	AR(4) in levels with constant
A2	AR(4) in levels with linear trend
A3	AR(4) in first differences with constant
A4	AR(4) in first differences with linear trend
A5	AR(4) with constant, pretested for unit root
A6	AR(4) with linear trend, pretested for unit root
A7	AR in levels with constant, AIC for lag length
A8	AR in levels with linear trend, AIC for lag length
A9	AR in first differences with constant, AIC for lag length
A10	AR in first differences with linear trend, AIC for lag length
A11	AR with constant, pretested for unit root, AIC for lag length
A12	AR with linear trend, pretested for unit root, AIC for lag length
A13	AR in levels with constant, BIC for lag length
A14	AR in levels with linear trend, BIC for lag length
A15	AR in first differences with constant, BIC for lag length
A16	AR in first differences with linear trend, BIC for lag length
A17	AR with constant, pretested for unit root, BIC for lag length
A18	AR with linear trend, pretested for unit root, BIC for lag length
B. Nonlinear models: Logistic smooth transition autoregressions (12 variants)	
L1	LSTAR(2) in levels, transition var. y_t
L2	LSTAR(2) in first differences, transition var. y_t
L3	LSTAR(2), pretested for unit root, transition var. y_t
L4	LSTAR(2) in levels, transition var. $y_t - y_{t-2}$
L5	LSTAR(2) in first differences, transition var. $y_t - y_{t-2}$
L6	LSTAR(2), pretested for unit root, transition var. $y_t - y_{t-2}$
L7	LSTAR in levels, AIC on transition var. and lag length
L8	LSTAR in first differences, AIC on transition var. and lag length
L9	LSTAR, pretested for unit root, AIC on transition var. and lag length
L10	LSTAR in levels, BIC on transition var. and lag length
L11	LSTAR in first differences, BIC on transition var. and lag length
L12	LSTAR, pretested for unit root, BIC on transition var. and lag length

Table 2: Number of factors used in backdating and the cumulated percentage of explained variance

Variable	Nr. of factors	Cum. Variance Prop.
YER	6	0.710
YED	6	0.762
CPI	4	0.731
EER	1	0.845
LTN	4	0.708
STN	5	0.724

Table 3: Best performing forecasting model/data variants

	$h = 1$			$h = 2$			$h = 4$		
Variable	Model	Data	MSFE	Model	Data	MSFE	Model	Data	MSFE
YER	A2[a]	FAC	0.858	A2	FAC	0.663	A2	FAC	0.406
	A14	FAC	0.906	A8[b]	FAC	0.687	A8[c]	FAC	0.430
	A2	AWM	0.985	A8[d]	AWM	0.944	L11[e]	FAC	0.900
YED	L8[f]	FAC	0.843	L8[g]	FAC	0.897	L4	AWM	0.667
	A1[h]	FAC	0.867	A1[i]	FAC	0.925	L7[j]	FAC	0.900
	L11[k]	FAC	0.934	L8[l]	AWM	0.933	L8[m]	FAC	1.003
CPI	L8[n]	FAC	0.390	L11[o]	AWM	0.523	L4	FAC	0.498
	L7	FAC	0.405	L10	AWM	0.558	L5[p]	FAC	0.723
	L11[q]	FAC	0.498	L10	FAC	0.694	L7	FAC	0.767
EER	A1	FAC	0.976	L4[r]	AWM	0.826	A1	FAC	0.867
	A7	FAC	0.985	A1[s]	FAC	0.928	A7	FAC	0.891
	A11	FAC	0.994	L4	FAC	0.933	A1[t]	AWM	1.000
LTN	A3[u]	FAC	0.828	A8[v]	AWM	0.780	A14	AWM	0.545
	A3[w]	AWM	0.843	A15[x]	AWM	0.788	A2	AWM	0.561
	L8[y]	FAC	0.848	A8[z]	FAC	0.803	A14	FAC	0.565
STN	L8[aa]	FAC	0.758	L3	FAC	0.520	L5	FAC	0.446
	A14	FAC	0.780	A3[bb]	AWM	0.611	A15	AWM	0.452
	A8	AWM	0.793	L2	FAC	0.623	A3[cc]	AWM	0.456

Note: AWM denotes forecasting model based on aggregated area-wide model data, FAC denotes model based on factor-backdating procedure. Entries in column 'MSFE' are MSFEs relative to benchmark model. The benchmark is an AR(4) with constant based on AWM data. Model names correspond to those from Table 1.

[a] A8 leads to same MSFE. [b] A14 leads to same MSFE. [c] A14 leads to same MSFE. [d] A14 leads to same MSFE. [e] L12 leads to same MSFE. [f] L9 leads to same MSFE. [h] A7 leads to same MSFE. [i] A7 and A13 lead to same MSFE. [j] L10 leads to same MSFE. [k] L12 leads to same MSFE. [l,m] L9 leads to same MSFE. [n] L9 leads to same MSFE. [o] L12 leads to same MSFE. [p] L6, L8 and L9 lead to same MSFE. [q] L12 leads to same MSFE. [r] L6 leads to same MSFE. [s] A7 leads to same MSFE. [t] A5 leads to same MSFE. [u] A5 and A6 lead to same MSFE. [v] A14 leads to same MSFE. [w] A5 and A6 lead to same MSFE. [x] A17 and A18 lead to same MSFE. [y] L9 leads to same MSFE. [z] A14 leads to same MSFE. [aa] L11 leads to same MSFE. [bb] A6, A7, A8, A9, A12, A15 and A18 lead to same MSFE. [cc] A6, A9 and A12 lead to same MSFE.

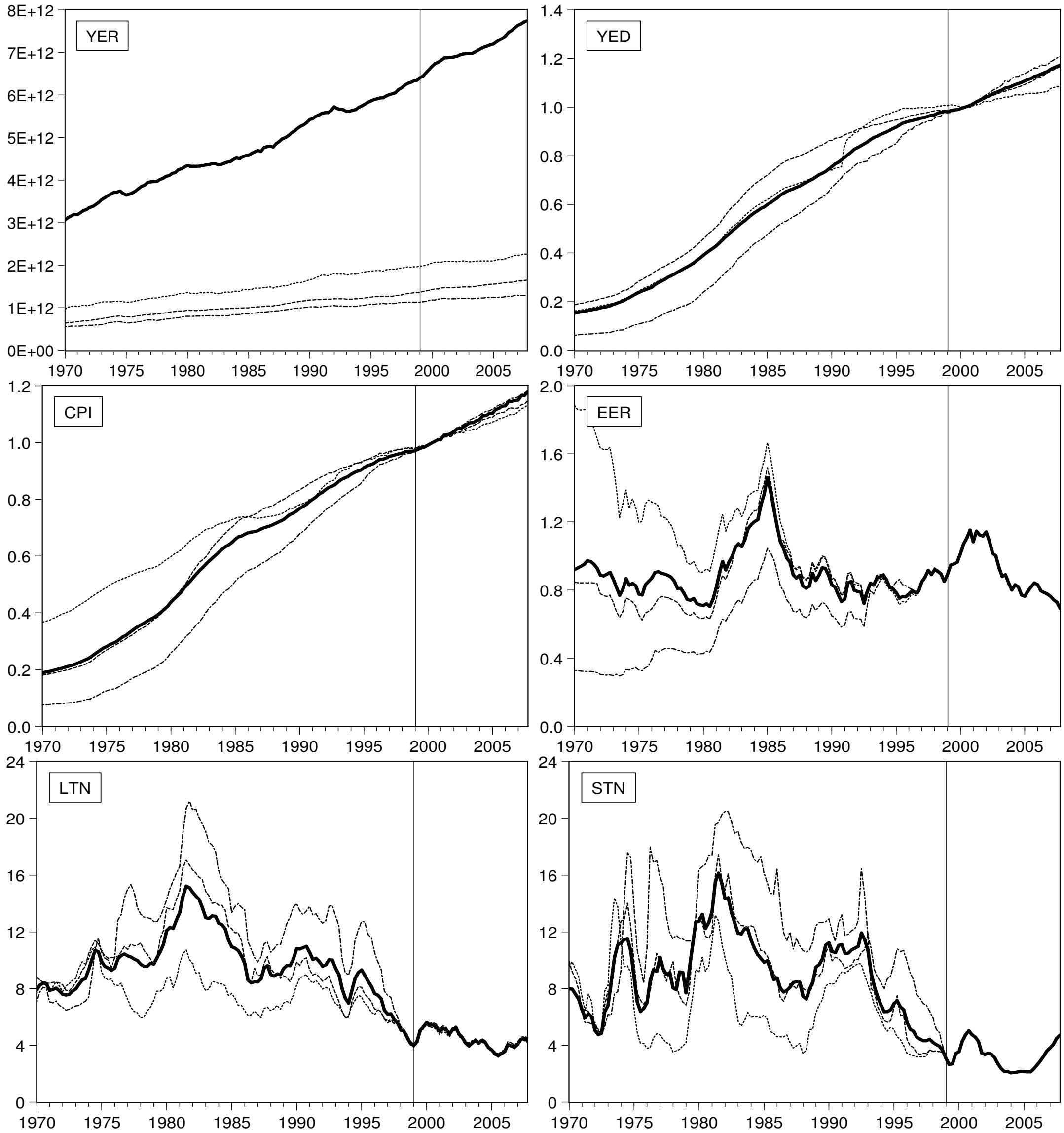

Figure 1: AWM Euro-area time series (solid lines) for real GDP (YER), the GDP deflator (YED), the consumer price index (CPI), the exchange rate (EER) and short- and long-term interest rates (STN and LTN). The dotted lines show time series plots of the corresponding variables for the three largest EMU member countries Germany, French and Italy.

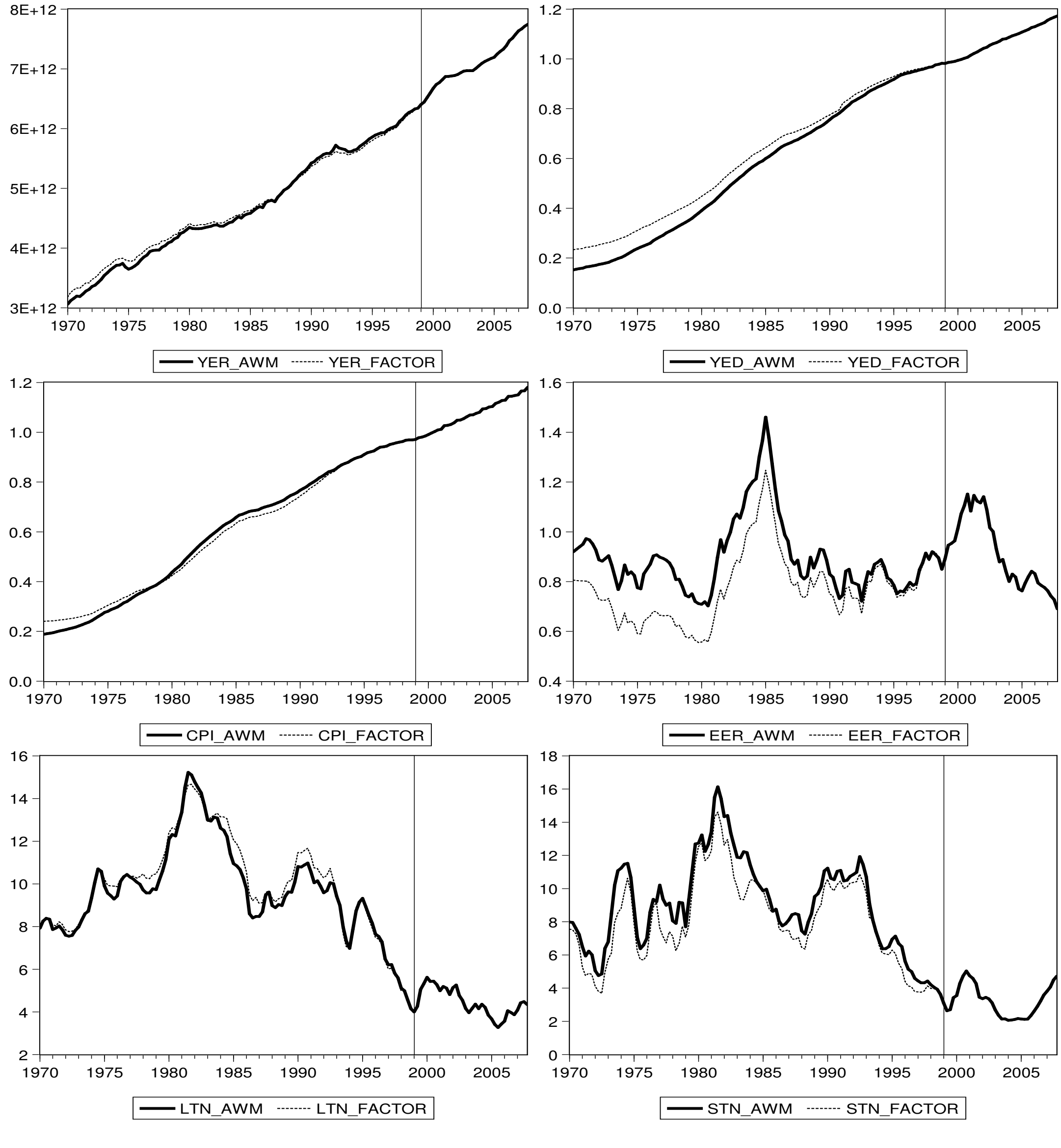

Figure 2: Euro-area time series for real GDP (YER), the GDP deflator (YED), the consumer price index (CPI), the exchange rate (EER) and short- and long-term interest rates (STN and LTN). Area-wide model series (solid lines) and factor-backdated time series (dotted lines).

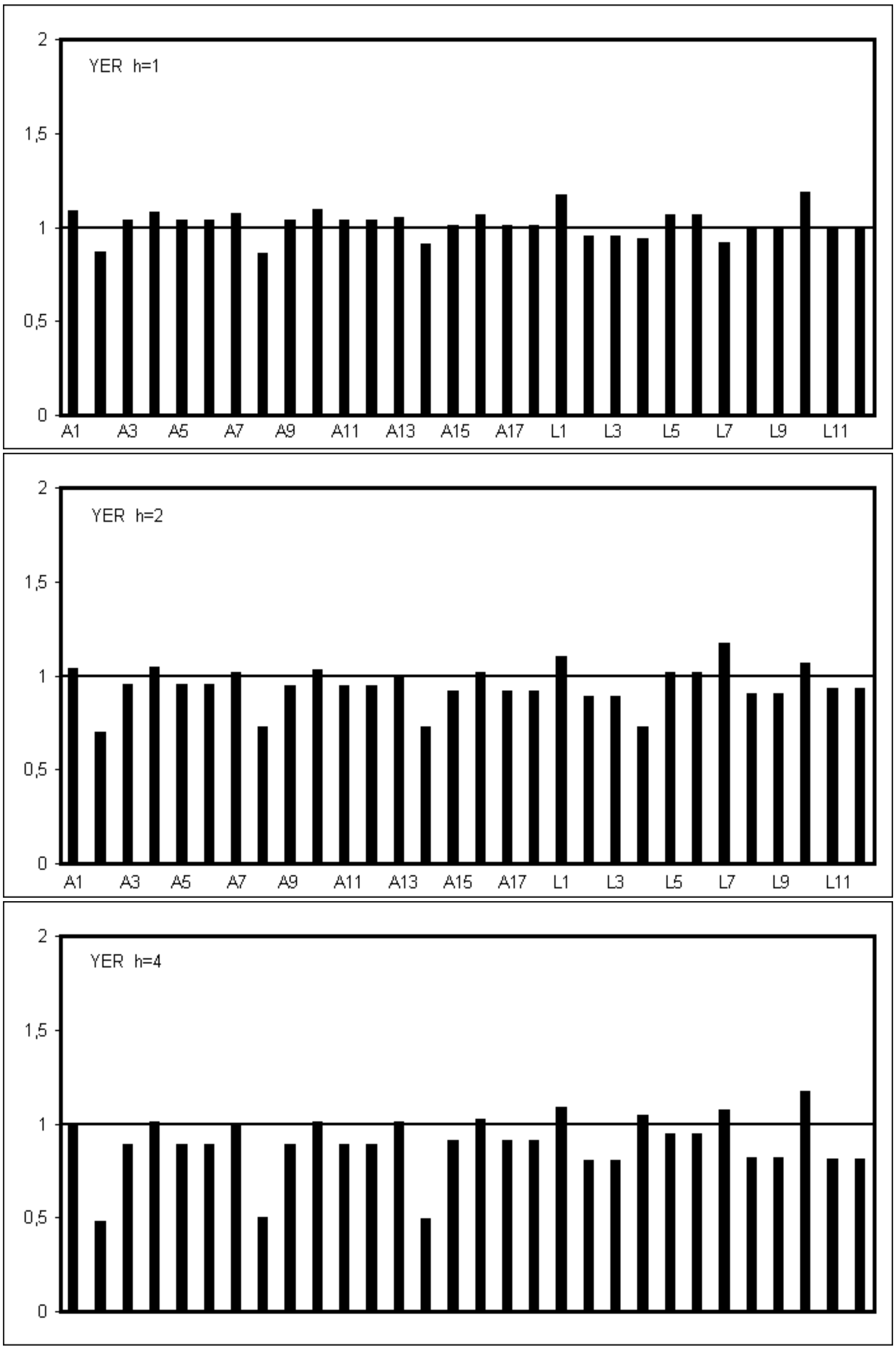

Figure 3: Results from forecasting comparison for real GDP (YER). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

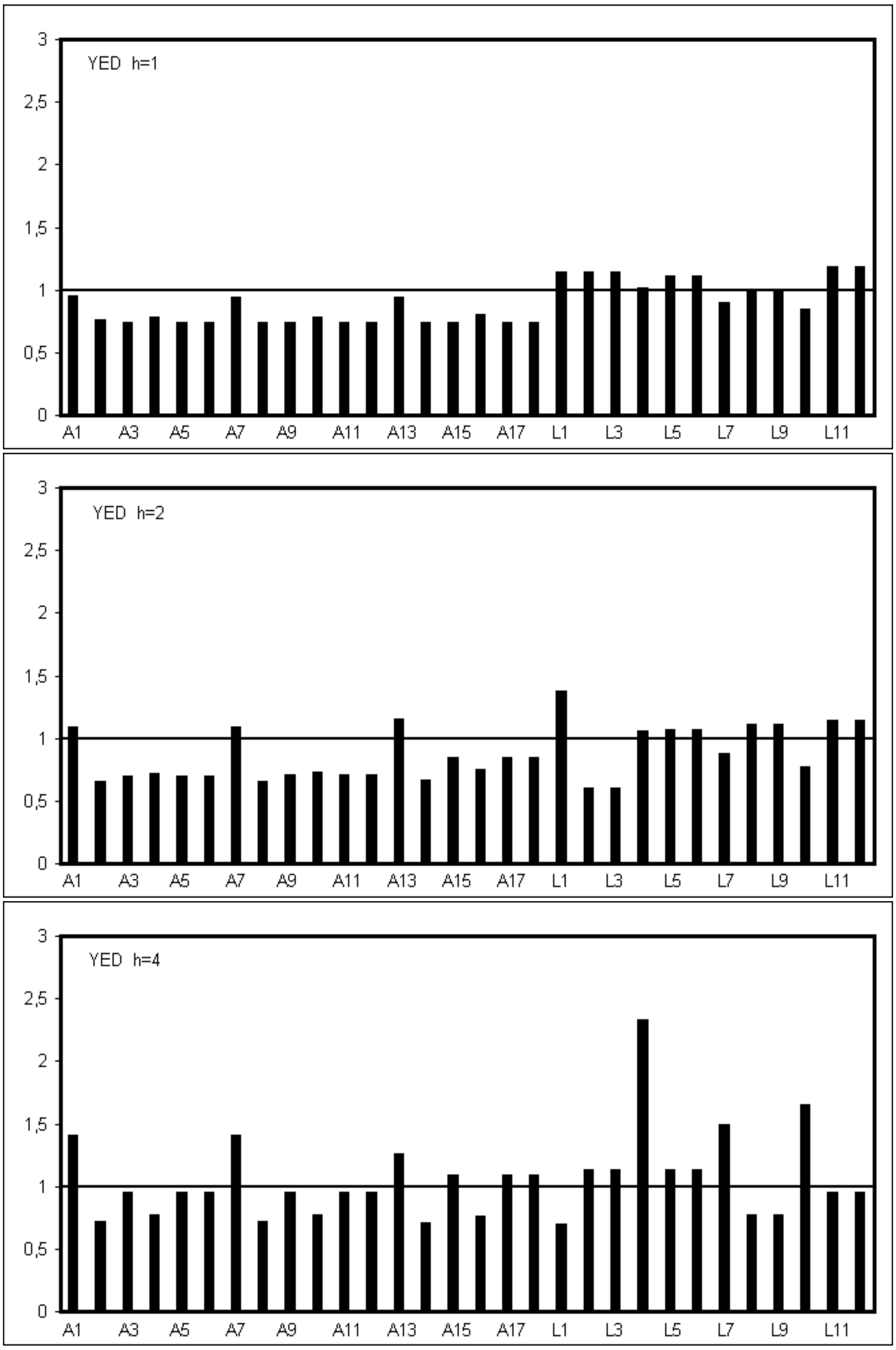

Figure 4: Results from forecasting comparison for GDP deflator (YED). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

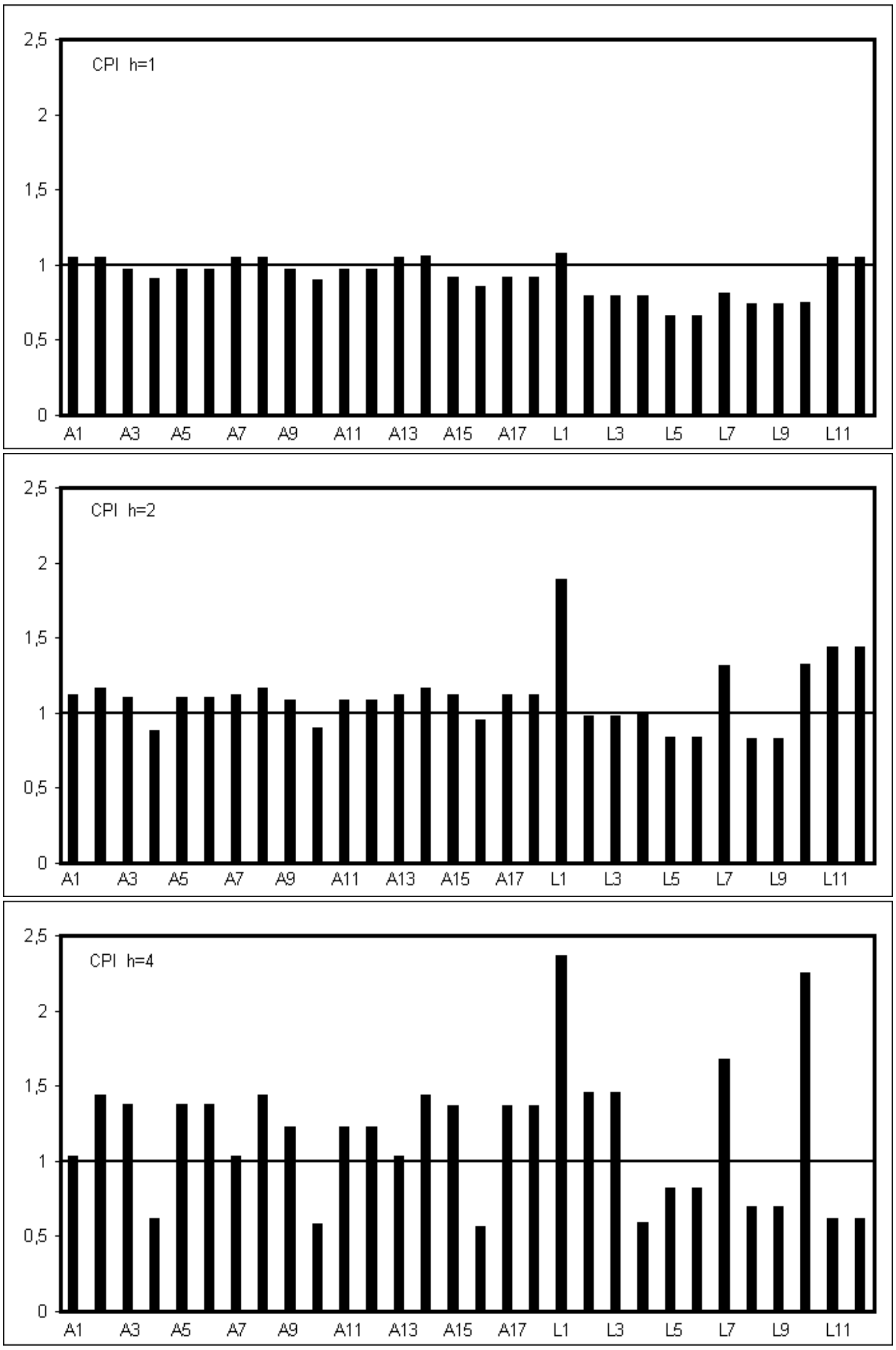

Figure 5: Results from forecasting comparison for the consumer price index (CPI). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

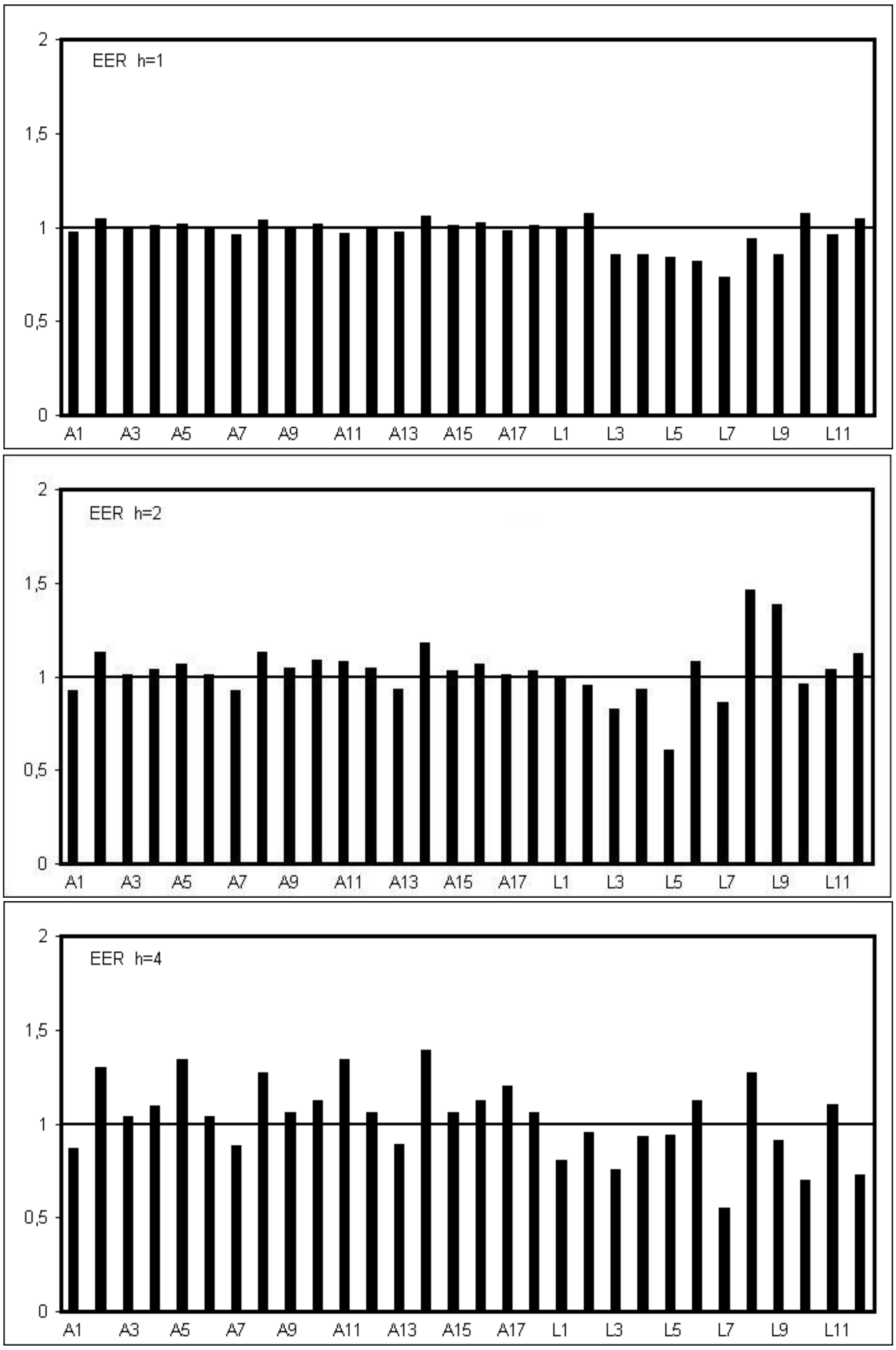

Figure 6: Results from forecasting comparison for the exchange rate (EER). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

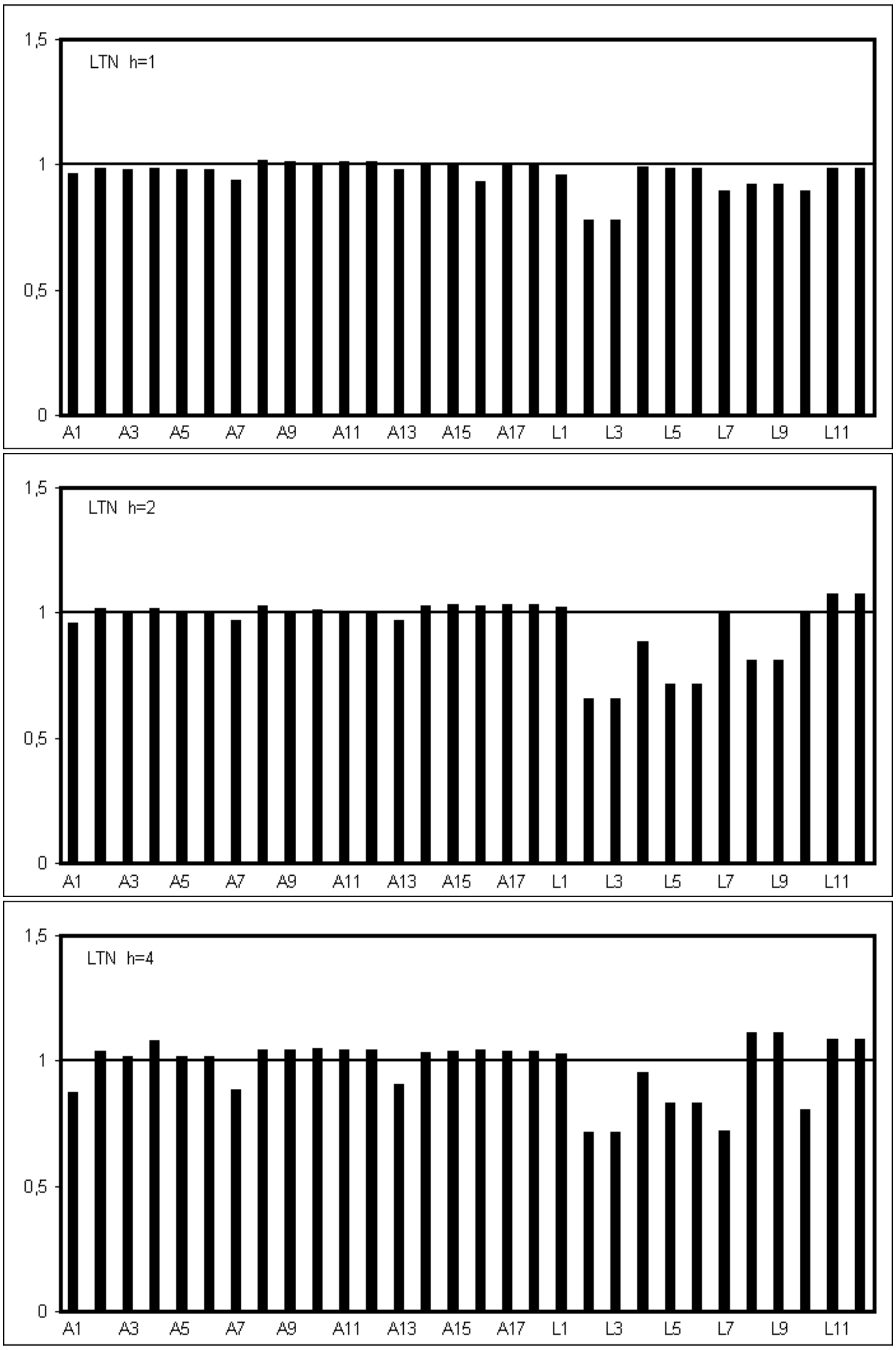

Figure 7: Results from forecasting comparison for long-term interest rate (LTN). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

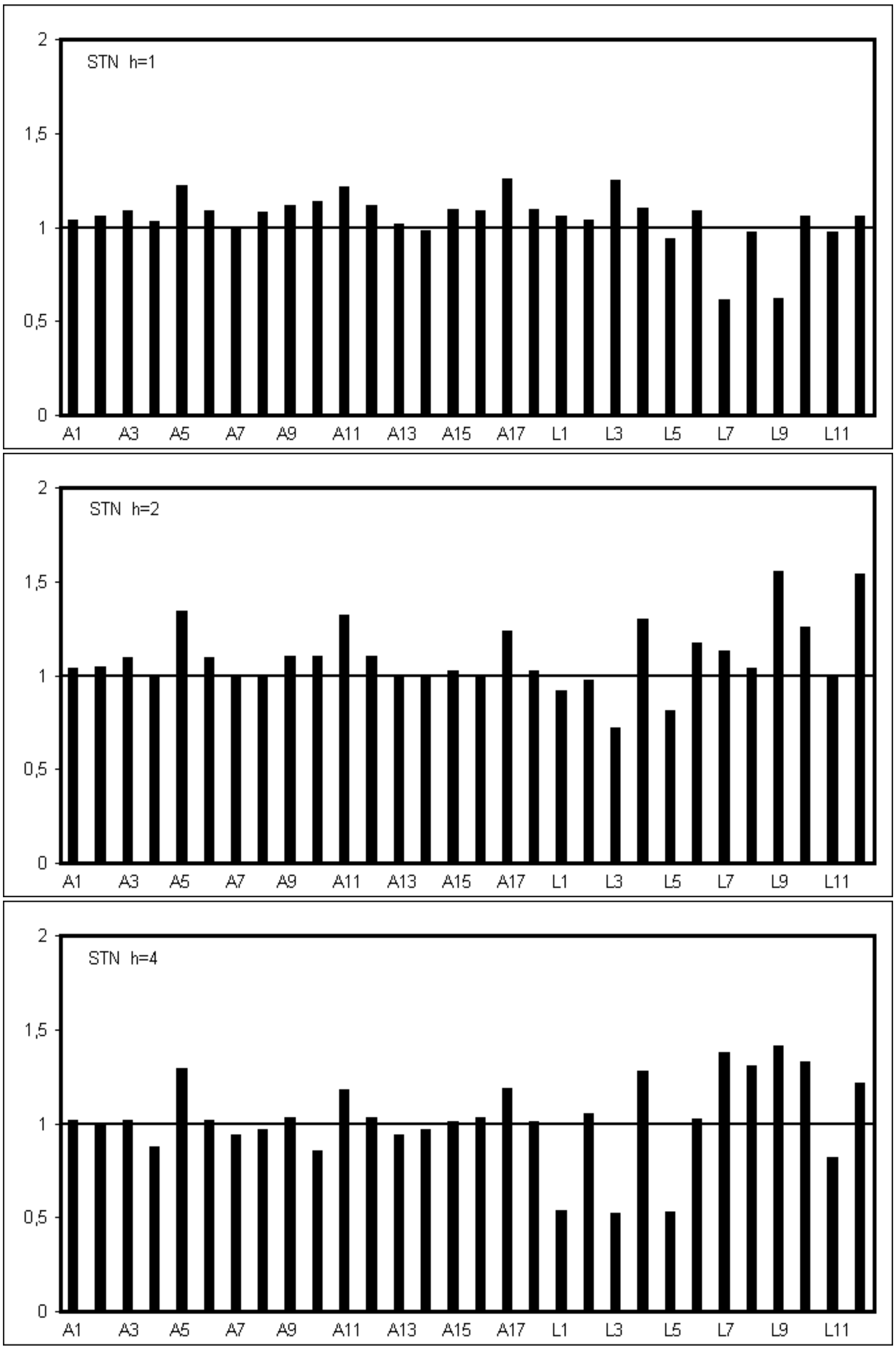

Figure 8: Results from forecasting comparison for short-term interest rate (STN). MSFEs of models using factor-backdated data relative to corresponding model with AWM data.

CMS Discussion Paper 2010-04

For a complete list of the Discussion Papers published by the CMS please visit http://www.uni-konstanz.de/cms.

01 "The Effect of Grade Retention on School Outcomes: An Application of Doubly Robust Estimation Method" by Selver Derya Uysal, January 2010

02 "Can we help finance professionals to predict the Euribor rate?" by Fabian Krüger, Frieder Mokinski and Winfried Pohlmeier, February 2010

03 "Bayesian Inference for Factor Structure Models via Gibbs Sampling" by Rémi Piatek, February 2010

04 "Forecasting Euro-Area Macroeconomic Variables Using a Factor Model Approach for Backdating" by Ralf Brüggemann and Jing Zeng, February 2010