

# Forecasting Aggregates with Stochastic Aggregation Weights

Helmut Lütkepohl

**Abstract** Despite the fact that the aggregation weights of many macroeconomic aggregates are time-varying, much of the literature on forecasting aggregates considers the case of fixed, time-invariant aggregation weights. In this study a framework for contemporaneous aggregation with possibly stochastic or time-varying weights is developed and different predictors for an aggregate are compared. An example based on European unemployment series is used to illustrate the virtue of the theoretical setup and the forecasting results.

**Key words:** Forecasting, stochastic aggregation, autoregression, moving average, vector autoregressive process

## 1 Introduction

Forecasting macroeconomic variables is a standard task of many applied economists. A number of economic research institutes and central banks publish macroeconomic forecasts regularly. Because most macroeconomic variables are contemporaneous aggregates it is not surprising that a huge literature exists on forecasting contemporaneously aggregated variables (see Lütkepohl (2010a) for a recent survey). Different predictors have been compared in this literature. Most results refer to linear aggregation with time-invariant aggregation weights.

On the other hand, in practice, time-varying aggregation weights are very common. Suppose, for example, that one wants to construct an unemployment rate series for the European Union (EU) by aggregating the unemployment rates of the individual EU countries. The overall unemployment rate,  $u_t^{EU}$ , is a weighted average of the individual unemployment rates,  $u_t^{(i)}$ ,

---

Helmut Lütkepohl  
Department of Economics, European University Institute  
Via della Piazzola 43, I-50133 Firenze, ITALY e-mail: [helmut.luetkepohl@eui.eu](mailto:helmut.luetkepohl@eui.eu)

$$u_t^{EU} = \sum_{i=1}^N w_{it} u_t^{(i)}. \quad (1)$$

Here  $N$  denotes the number of member states and the aggregation weight  $w_{it}$  is related to the relative size of the workforce of country  $i$ . Because the relative country sizes change over time, the weights in (1) also have to be time-varying. In fact, the number of EU countries has changed over the years and is meant to change further in the future when new member states enter the EU. Hence, one may use for  $N$  the maximum number of states and assign zero weights to those countries which were not members at a particular time  $t$ . Even for periods with a constant number of member states the weights are related to the workforce in the different member states which is best viewed as stochastic. Hence, the weights are random variables. In other words, the EU unemployment rate is a contemporaneous aggregate with stochastic aggregation weights.

In this study a general framework is presented for comparing alternative forecasts for nonlinear aggregates including aggregates with stochastic weights. The basic idea is adapted from Hendry and Hubrich (2010) who propose to incorporate disaggregate information in forecasting an aggregate by using a system of variables comprised of the aggregate and some disaggregate series. An empirical forecast comparison for European unemployment variables is used to illustrate the theoretical results. The study summarizes results from Lütkepohl (2010b) where more details and references to related literature can be found.

The study is organized as follows. In the next section the basic setup is presented. Theoretical results on alternative predictors for nonlinear aggregates with possibly stochastic weights are discussed in Section 3. Empirical forecast comparisons are discussed in Section 4. Conclusions are drawn in Section 5.

The following general notation is used. The lag operator is denoted by  $L$ , that is, for a stochastic process or time series  $y_t$ ,  $Ly_t = y_{t-1}$ . The normal distribution with mean (vector)  $\mu$  and variance (covariance matrix)  $\Sigma$  is denoted by  $\mathcal{N}(\mu, \Sigma)$ . The sets of integers and positive integers are signified as  $\mathbb{Z}$  and  $\mathbb{N}$ , respectively. The  $(K \times K)$  identity matrix is denoted by  $I_K$ . Moreover, the following abbreviations are used: AR for autoregressive or autoregressive process, VAR for vector autoregressive, MA for moving average, ARCH for autoregressive conditional heteroskedasticity, iid for independently identically distributed, OLS for ordinary least squares, GLS for generalized least squares, DGP for data generating process, MSE for mean squared error, RMSE for root mean squared error, AIC for the Akaike information criterion and SC for the Schwarz-Rissanen information criterion (see, e.g., Lütkepohl (2005, Section 4.3) for precise definitions and discussion of these criteria).

## 2 Aggregates

Let  $y_t = (y_{1t}, \dots, y_{Kt})'$  be the vector of disaggregate component series and the aggregate of interest is  $a_t = f_t(y_t)$ . Here  $f_t(\cdot)$  is a scalar aggregation function which may

be time-varying. In fact, it is straightforward to extend the discussion to a vector process  $a_t$ . The assumption of a univariate process is made for clarity and notational convenience. A typical situation is that  $f_t(y_t) = w_t' y_t$ , where  $w_t = (w_{1t}, \dots, w_{Kt})'$  is a vector of stochastic weights.

The process  $a_t$  is assumed to be (covariance) stationary with AR and MA representations,

$$a_t = \sum_{i=1}^{\infty} \theta_i a_{t-i} + v_t, \quad t \in \mathbb{Z}, \quad (2)$$

and

$$a_t = \sum_{i=0}^{\infty} \psi_i v_{t-i}, \quad t \in \mathbb{Z}, \quad (3)$$

respectively, where  $v_t$  is a martingale difference white noise sequence with respect to  $\mathcal{A}_t = \{a_t, a_{t-1}, \dots\}$  with time-invariant variance  $\sigma_v^2$ . The martingale difference property implies that the conditional mean  $E(v_t | a_{t-1}, a_{t-2}, \dots) = 0$ .

Notice that for an aggregate  $a_t = w_t' y_t$  covariance stationarity holds under quite general conditions for  $w_t$  and  $y_t$ . For example, if the two processes are both covariance stationary and independent.

Suppose that  $y_t^*$  is an  $M$ -dimensional (covariance) stationary stochastic process related to  $y_t$ . It may be equal to  $y_t$  or it may consist of a subset of the components of  $y_t$ . In fact, it may also contain transformations of the components of  $y_t$  or, more generally, any variables that are related to  $y_t$ . It is assumed that the joint process  $x_t = (a_t, y_t^{*'})'$  is stationary with VAR and MA representations

$$x_t = \sum_{i=1}^{\infty} A_i x_{t-i} + e_t, \quad t \in \mathbb{Z}, \quad (4)$$

and

$$x_t = \sum_{i=0}^{\infty} \Xi_i e_{t-i}, \quad t \in \mathbb{Z}, \quad (5)$$

respectively, where the  $A_i$ 's and  $\Xi_i$ 's are  $((M+1) \times (M+1))$  dimensional coefficient matrices with  $\Xi_0 = I_{M+1}$  and  $e_t$  is a  $(M+1)$ -dimensional martingale difference with respect to  $\mathcal{X}_t = \{x_t, x_{t-1}, \dots\}$  with time-invariant, nonsingular covariance matrix  $\Sigma_e$ .

Assuming nonsingularity of the covariance matrix  $\Sigma_e$  means that none of the components of  $x_t$  is a linear combination of the remaining components. Thus, if  $a_t = w_t' y_t$  is an aggregate of  $y_t$  with fixed, time-invariant weights,  $y_t^* = y_t$  is not possible. However, as long as at least one component series of  $y_t$  is missing from  $y_t^*$ , time-invariant aggregation is covered. Thus, the present framework can be used for studying forecasts of aggregates with fixed weights  $w$  by including the aggregate and, say, the last  $K-1$  components of  $y_t$  in  $y_t^*$ .

If the aggregation weights are stochastic, considering the aggregate  $a_t$  in addition to the disaggregate component series  $y_t$ , that is, choosing  $y_t^* = y_t$  leads to an extension of the information set because the information in the process  $w_t$  is also incorporated. These considerations imply that the aforementioned framework is useful not only for studying stochastic aggregation but any form of aggregation which implies a nonsingular joint process  $x_t$ , e.g., whenever  $e_t$  has a nonsingular covariance matrix. The latter condition may also be satisfied if  $a_t$  is some other nonlinear aggregate of  $y_t$ .

The framework is useful even if some of the disaggregate component series are not available or if they are not known precisely. In that case one may just include all available component series in  $y_t^*$ . Also one may include series in  $y_t^*$  which are potentially useful indicators for the aggregate. For example, one may include factors which are computed from a large set of disaggregate component series. Moreover, the framework can be used if the weights are deterministic but still have changed over time, e.g., due to a structural break or a new member in some union.

Note that deterministic terms have been deleted from the DGPs for convenience in the present setup because they are not of much interest for the theoretical comparison of predictors. Since future values of deterministic terms are known, they do not contribute to forecast uncertainty. Excluding them in the theoretical analysis is therefore not restrictive. Of course, they have to be included when models are fitted to actual data.

The present framework is limited to stationary processes. If AR unit roots are the only source of nonstationarity, this feature can be accounted for by considering differenced variables. Most of the results discussed in the following will also hold for integrated processes. I will comment on this issue later. For the initial comparison of different predictors in the next section, the stationarity assumption is helpful because it simplifies the exposition and analysis.

Assuming that the residuals of the DGPs under consideration are martingale differences may exclude potentially interesting nonlinearities. It is made for convenience because it implies that optimal predictors are easy to determine. As will be seen in the next section, the predictors under scrutiny will still be best *linear* predictors if the white noise processes in the DGPs are at least serially uncorrelated and not necessarily martingale differences. In that case, the results on the relative efficiencies of the predictors are maintained.

### 3 The Predictors

If forecasts of  $a_t$  are desired, there are at least two obvious predictors. Suppose a forecast  $h$  periods ahead (an  $h$ -step forecast) is of interest in period  $\tau$ . The optimal, minimum mean squared error (MSE) forecast based on present and past values of  $a_t$  is the conditional expectation,

$$a_{\tau+h|\tau} = E(a_{\tau+h}|a_{\tau}, a_{\tau-1}, \dots) = \sum_{i=1}^{\infty} \theta_i a_{\tau+h-i|\tau} = \sum_{i=h}^{\infty} \psi_i v_{\tau+h-i}, \quad (6)$$

where  $a_{\tau+h-i|\tau} = a_{\tau+h-i}$  for  $i \geq h$ . Thus, the AR form can be used for computing forecasts recursively for  $h = 1, 2, \dots$ , while the last expression on the right-hand side of (6) is useful for assessing the forecast error. The corresponding quantity is

$$a_{\tau+h} - a_{\tau+h|\tau} = \sum_{i=0}^{h-1} \psi_i v_{\tau+h-i} \sim (0, \sigma_a^2(h)), \quad (7)$$

where

$$\sigma_a^2(h) = \sigma_v^2 \sum_{j=0}^{h-1} \psi_j^2 \quad (8)$$

is the forecast error variance which is also the forecast MSE. The fact that the conditional expectation can be expressed as in (6) is a consequence of the assumption that  $v_t$  is a white noise martingale difference sequence.

The second obvious predictor for  $a_t$  is based on the full process  $x_t$ . Let

$$x_{\tau+h|\tau} = E(x_{\tau+h}|x_{\tau}, x_{\tau-1}, \dots) = \sum_{i=1}^{\infty} A_i x_{\tau+h-i|\tau} = \sum_{i=h}^{\infty} \Xi_i e_{\tau+h-i} \quad (9)$$

be the optimal predictor based on present and past values of  $x_t$ . Its forecast error covariance or MSE matrix is

$$\Sigma_x(h) = E[(x_{\tau+h} - x_{\tau+h|\tau})(x_{\tau+h} - x_{\tau+h|\tau})'] = \sum_{j=0}^{h-1} \Xi_j \Sigma_e \Xi_j'. \quad (10)$$

The first component of  $x_{\tau+h|\tau}$  is an  $h$ -step forecast of  $a_{\tau+h}$  which will be denoted by  $a_{\tau+h|\tau}^o$  in the following to distinguish it from  $a_{\tau+h|\tau}$ . The corresponding forecast error has zero mean and variance or MSE

$$\sigma_o^2(h) = \mathbf{i}' \Sigma_x(h) \mathbf{i}, \quad (11)$$

where  $\mathbf{i} = (1, 0, \dots, 0)'$  is a  $(M+1)$ -dimensional vector.

A well-known forecasting result for linearly transformed vectors implies that

$$\sigma_o^2(h) \leq \sigma_a^2(h) \quad (12)$$

(see, e.g., Lütkepohl (1987)). Moreover, stationarity of  $a_t$  implies that  $a_{\tau+h|\tau}^o$  and  $a_{\tau+h|\tau}$  become identical for  $h \rightarrow \infty$  and, hence,

$$\lim_{h \rightarrow \infty} \sigma_o^2(h) = \lim_{h \rightarrow \infty} \sigma_a^2(h) = E(a_t^2). \quad (13)$$

Thus, the two predictors are equally efficient for long-term forecasting.

In fact, under rather special conditions they may also be equal for short forecast horizons. This may hold, for instance, if the MA operator of  $y_t$  is diagonal and the individual component series of  $y_t$  have homogeneous dynamics. For practical purposes, using the predictor based on the aggregated process directly may be justified even if predictor equality holds only approximately. Even that requires a very special DGP of the disaggregate process  $y_t$ , however, and, hence, for the case of a time-invariant aggregation scheme, gains from using disaggregated information for short-term forecasting seem likely in typical situations. For long-term forecasts of stationary variables the disaggregate information is not helpful, however, because  $a_{\tau+h|\tau}^o$  and  $a_{\tau+h|\tau}$  will be similar or identical for large  $h$ .

For the general case, where  $a_t = f_t(y_t)$ , the discussion in Section 4.2.4 in Lütkepohl (1987) shows that equality of the two predictors,  $a_{\tau+h|\tau}^o = a_{\tau+h|\tau}$  for all  $h \in \mathbb{N}$ , holds if and only if  $y_t^*$  does not Granger-cause  $a_t$  or, in other words, if there are no lagged  $y_t^*$ 's in the  $a_t$  equation in (4). For instance, for the example considered in the introduction, the forecast based on the disaggregate information is generally superior to  $a_{\tau+h|\tau}$  because the components of  $y_t$  are included in  $a_t$  and, hence, will generally enter the  $a_t$  equation in the model for  $x_t$ .

If the variables of interest have unit roots and are, hence, integrated, these general results will still hold because stationarity can be achieved by differencing. Since forecasts for the levels of the variables can be obtained directly from forecasts of the changes it is in fact enough to focus on stationary processes for deriving theoretical results which also hold for MA processes with unit roots. MA unit roots are permitted, for example, in deriving the superiority of the forecast based on disaggregate components, that is, the result in (12). Thus, even if all components are differenced individually without regard of possible cointegration, the previous theoretical analysis remains valid. For practical purposes one may still work with the levels VAR version of the process and avoid the detour via the differences. What does not hold for integrated processes is the convergence of the forecast MSEs in (13). In other words, even for long-term forecasting,  $a_{\tau+h|\tau}^o$  may be superior to  $a_{\tau+h|\tau}$  if  $a_t$  has a unit root.

Hendry and Hubrich (2010) consider this framework although they assume time-invariant aggregation weights throughout much of their theoretical analysis. They consider the predictors  $a_{\tau+h|\tau}^o$  and  $a_{\tau+h|\tau}$  for U.S. inflation. Note, however, that the main focus of Hendry and Hubrich (2010) is an analysis of the impact of various types of model deficiencies.

Although the previous results suggest that theoretically a forecast of the aggregate should always be based on the disaggregate process  $x_t$  because the components of  $y_t$  are part of the aggregate and, hence, are related to  $a_t$  and supposedly should have a Granger-causal relation to  $a_t$ , the situation is more difficult in practice, as pointed out by Hendry and Hubrich (2010). The theoretical results refer to correctly specified models with known coefficients. In practice, models for the DGPs involved have to be estimated and specified on the basis of the available sample information. As shown already by Lütkepohl (1987), estimation and specification uncertainty can reverse the ranking of the MSEs of the two predictors. In fact, if a

large set of component series is aggregated, specifying a joint multivariate model for the components and the aggregate may not be feasible because of degrees of freedom limitations. Even if just a moderate number of components is considered, fitting a multivariate model may involve estimation of many more parameters and, hence, a much larger estimation uncertainty than fitting a univariate model to the aggregate only. Thereby the forecast based on the disaggregate information may become inferior to the one based on the aggregate directly. Therefore Hendry and Hubrich (2010) propose to limit the number of components included in the multivariate model. For example, for an inflation forecast one may include only the most important sub-index information rather than considering all sectoral indices. Alternatively, one may summarize the information of a large number of components in a small number of factors which are included in the multivariate model.

I have explored the small sample implications of including disaggregate information in the forecast for an aggregate based on stochastic weights with a Monte Carlo study. The results show that, depending on the properties of the DGP, using disaggregate information can be valuable for improving forecast precision for an aggregate. On the other hand, if the disaggregate components are rather homogenous, for example, adding many of them can be detrimental to the forecast accuracy. Thus, to obtain actual gains in forecast RMSE over a univariate forecast for the aggregate, limiting the parameter space by considering a small number of disaggregate components and/or a small number of lags is advantageous if samples of typical sizes in macroeconomics are available. If relatively large samples are available and the disaggregate components are heterogeneous, adding more of them to the model may be a useful strategy for improving forecasts of the aggregate. Finally, depending on the persistence of the DGP and the forecast horizon, little or no gains can be expected from using disaggregate information for long-term forecasting of stationary DGPs.

## 4 Empirical Forecast Comparison

I use euro-area unemployment rates to explore the empirical relevance of the previous results. The dataset consists of monthly, seasonally adjusted unemployment rates for the euro-area ( $a_t$ ) and 10 individual euro-area countries for the period 1995M1-2008M12. Thus, given the relatively large number of component series, I consider only a subset of the 16 EMU countries for which the aggregate  $a_t$  is computed. A number of smaller countries have been eliminated because their relative weight in the overall unemployment rate is small. Details of the data are given in the Appendix.

I have fitted AR and VAR models to the overall euro-area unemployment rate and subsets of the available country series. Clearly including all 11 series in a full VAR would have severely reduced the degrees of freedom. One of the questions of interest in the following is how many series should be included in a model that accounts for disaggregate information. Hence, the forecast performance of systems of different dimensions is explored. All AR and VAR models are based on the levels

unemployment rate series although unit roots are not rejected by standard tests for some country rates. As I have argued in Section 3, unit roots are no problem if AR and VAR models are used for forecasting. Whether they can be utilized to improve forecasts is not the focus of this study. If not accounting explicitly for unit roots reduces forecast efficiency, it will affect all predictors although perhaps not to the same extent.

Because the simulation results suggest that parsimonious models may be beneficial for forecasting, I fit full and subset AR and VAR models with an intercept. The maximum order is set at  $p_{\max} = 6$  because this order appears to be large enough to capture much if not all of the serial dependence structure for the systems considered. When full AR and VAR models are considered the actual model order is chosen by a model selection criterion. AIC and SC are used for this purpose. Estimation of full models is done by OLS. When subset models are used, the subset restrictions are determined by a sequential elimination strategy for each equation separately when multivariate models are considered. The sequential procedure proceeds by estimating an equation by OLS and eliminating the regressor with the smallest absolute  $t$ -value. Then the restricted model is estimated and the next regressor with the smallest absolute  $t$ -ratio is eliminated until all absolute  $t$ -ratios are greater than 1.96. Thus, a 5% significance level is used as a threshold for eliminating regressors in the subset models. Once all zero restrictions for all the equations of a given model are determined, the restricted model is reestimated by GLS with the OLS covariance estimator of a full VAR model with order  $p_{\max}$  (see, Section 5.2.2 of Lütkepohl (2005) for details on the GLS procedure).

Following Hendry and Hubrich (2010), I also consider models with factors which are computed as principle components from all 10 disaggregate country data. They present another way of obtaining parsimonious models. The factors are computed on the basis of the sample available for model specification and estimation. They are recomputed whenever the sample is extended by an observation. Then VAR models are fitted for the aggregate and one or more factors. Again full and subset VAR models are considered for systems including factors. The previously described specification procedures are used here as well.

Forecasts are computed based on models fitted to increasingly larger samples. The last four years are set aside for the forecast comparison. Thus, the first set of models is based on samples from 1995M1-2004M12 (i.e., 120 observations). Based on these models 1- to 12-step forecasts are produced for the period 2005M1-2005M12. Then one observation is added to the sample and model specification and estimation are repeated for all models. Thereby 48 1-step, 47 2-step etc. out-of-sample forecasts are produced and used for computing RMSEs. Notice that whenever the sample is increased by a new observation, a full new specification and estimation is performed and also new factors are computed. It is perhaps worth emphasizing that multi-step forecasts are computed recursively as in Section 3 (see Equations (6) and (9)).

Relative forecast RMSEs for a number of different models and forecast horizons  $h = 1, 6$  and 12 are presented in Table 1, where the univariate AR model specified with the parsimonious SC is used as a benchmark. In other words, numbers smaller

than one in the table indicate that the corresponding model produced a smaller RMSE than the univariate full AR model for  $a_t$  with lag order selected by SC. The following observations emerge from Table 1:

**Table 1** Forecast RMSEs for Euro-Area Unemployment Relative to Univariate AR Model Selected by SC

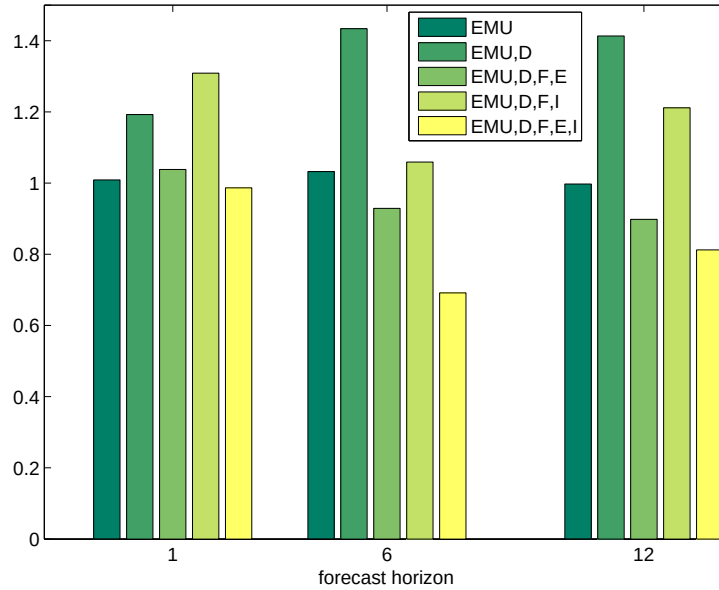
| Model type    | Model                          | forecast horizon |        |         |
|---------------|--------------------------------|------------------|--------|---------|
|               |                                | 1-step           | 6-step | 12-step |
| full AR/VAR   | SC-VAR <sup>EMU,D</sup>        | 0.9504           | 0.8227 | 0.6737  |
|               | SC-VAR <sup>EMU,D,F,E</sup>    | 0.9773           | 0.7716 | 0.6966  |
|               | SC-VAR <sup>EMU,D,F,I</sup>    | 0.9847           | 0.8629 | 0.8801  |
|               | SC-VAR <sup>EMU,D,F,E,I</sup>  | 1.2519           | 1.1613 | 1.1370  |
|               | SC-VAR <sup>EMU,f1</sup>       | 0.9466           | 0.8006 | 0.6818  |
|               | SC-VAR <sup>EMU,f1,f2</sup>    | 1.1357           | 1.3319 | 1.5271  |
|               | AIC-AR <sup>EMU</sup>          | 1.0000           | 1.0000 | 1.0000  |
|               | AIC-VAR <sup>EMU,D</sup>       | 0.9227           | 0.8080 | 0.6508  |
|               | AIC-VAR <sup>EMU,D,F,E</sup>   | 0.9293           | 0.7144 | 0.5847  |
|               | AIC-VAR <sup>EMU,D,F,I</sup>   | 0.9627           | 0.8006 | 0.7905  |
|               | AIC-VAR <sup>EMU,D,F,E,I</sup> | 1.0799           | 0.9854 | 0.9610  |
|               | AIC-VAR <sup>EMU,f1</sup>      | 0.9321           | 0.7949 | 0.6829  |
|               | AIC-VAR <sup>EMU,f1,f2</sup>   | 1.0727           | 1.1313 | 1.3272  |
|               | AIC-AR <sup>EMU</sup>          | 1.0090           | 1.0324 | 0.9975  |
| subset AR/VAR | VAR <sup>EMU,D</sup>           | 1.1334           | 1.1796 | 0.9522  |
|               | VAR <sup>EMU,D,F,E</sup>       | 1.0145           | 0.7167 | 0.6257  |
|               | VAR <sup>EMU,D,F,I</sup>       | 1.2888           | 0.9137 | 1.0663  |
|               | VAR <sup>EMU,D,F,E,I</sup>     | 1.2351           | 0.8028 | 0.9235  |
|               | VAR <sup>EMU,f1</sup>          | 1.0505           | 0.9154 | 1.0643  |
|               | VAR <sup>EMU,f1,f2</sup>       | 1.2047           | 1.0526 | 1.2614  |

*Notes:* Total sample period: 1995M1-2008M12; forecast period: 2005M1-2008M12; maximum lag order  $p_{\max} = 6$ ; EMU – European Monetary Union (euro-area), D – Germany, F – France, E – Spain, I – Italy; f1 – first factor, f2 – second factor.

1. Univariate models for the aggregate unemployment rate based on AIC and SC produce the same RMSEs, that is, AIC and SC select the same AR orders (all relative RMSEs of the AIC-AR model are 1).
2. Adding disaggregate individual country variables in the full VAR models helps reducing the RMSEs up to a certain point. The bivariate full VAR models with the aggregate EMU rate and the German rate produce substantially better forecasts than the univariate models. When more variables are added, RMSEs tend to go up. For example, the five-dimensional model comprised of the EMU, German, French, Spanish and Italian rates produce substantially larger RMSEs for all three horizons than the bivariate models.
3. Subset VAR models do not improve forecasts of univariate models and low-dimensional systems. They produce lower RMSEs than the corresponding full VAR models in some cases for 4- and 5-dimensional systems. Hence, parsimony considerations become important when the number of component series

increases. This result can also be seen in Figure 1, where the RMSEs of subset VARs relative to the corresponding full VARs are shown. The subset models have a clear advantage only for the five-dimensional system.

4. The previous observation suggests that factor models may be useful to include information from a larger number of individual country variables. This is true only partially, however. Bivariate models with one factor produce reasonably small RMSEs when full VAR models are used. Adding a second factor leads to larger RMSEs across all models. In fact, the two-factor models are substantially inferior to the benchmark models. Hence, they do not even improve on direct univariate forecasts of the aggregate. Generally, specifying factor models with subset procedures leads to poor forecasts.



**Fig. 1** RMSEs of subset AR/VAR models for unemployment relative to full SC-AR/VAR models.

In summary, the forecasting results for the euro-area unemployment rate are in line with the theoretical and simulation results. They confirm that disaggregate information is valuable for improving forecasts as long as the disaggregate models do not get very large. If too much disaggregate information is included in the model, the lack of parsimony may become an obstacle at the specification and estimation stage and the resulting models may lead to substantially inferior forecasts for the aggregate.

## 5 Conclusions

In this study a framework is presented for analyzing forecasts for aggregated variables obtained by nonlinear and possibly stochastic aggregation. Such aggregates are common in economics. The main advantage of the proposed framework is that the precise form of aggregation and in particular the aggregation weights may be unknown, as is often the case in practice. Using the proposed setup, forecasts based on the aggregate directly and forecasts which take into account at least some of the disaggregate information are compared. It is found that theoretically forecasts that take into account disaggregate information will generally lead to more efficient forecasts in terms of MSE. In practice, including too many disaggregate components can lead to efficiency losses because specification and estimation uncertainty will increase with the number of disaggregate components in the model used for forecasting. Hence, the additional information content in further disaggregate components may be insufficient to compensate for increases in specification and estimation uncertainty.

An example based on euro-area unemployment rates is used to illustrate the results. It is found that using disaggregate information from the individual member states may result in forecast improvements if the number of component series is small and parsimonious models are fitted.

These results raise questions for further research. In particular, it may be of interest to determine the optimal level of disaggregation. Given the results of this paper, it will depend on the DGP of the variables of interest and, hence, optimal disaggregation is ultimately an empirical question. Still, it may be possible to find characteristics of component series which give an indication that they may be helpful in a model for the aggregate. Such questions are left for future research.

**Acknowledgements** This paper is an abbreviated summary of Lütkepohl (2010b) where more details can be found. I thank Ralf Brüggemann, Andrea Silvestrini and Giacomo Sbrana for commenting on an earlier draft of the paper.

## Appendix. Data and Their Sources

All data were extracted from the Statistical Data Warehouse of the European Central Bank. The series are standardized unemployment series as collected by Eurostat. All ages and males and females are included. The series are monthly frequency. They are seasonally adjusted by Eurostat but not working day adjusted. The total sample period is from January 1995 to December 2008. The following eleven series are included in the panel: Euro-area 16 (fixed composition), Austria, Belgium, Germany, Spain, Finland, France, Ireland, Italy, Netherlands, Portugal.

## References

- Hendry, D. F. and Hubrich, K. (2010). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate, *Journal of Business & Economic Statistics*.
- Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.
- Lütkepohl, H. (2010a). Forecasting aggregated time series variables - a survey, *Journal of Business Cycle Measurement and Analysis*.
- Lütkepohl, H. (2010b). Forecasting nonlinear aggregates and aggregates with time-varying weights, *Working Paper EUI ECO 2010*, European University Institute.