Forecasting Aggregates by Disaggregates

David F. Hendry and Kirstin Hubrich*
Department of Economics, Oxford University
and Research Department, European Central Bank.

January, 2005

Abstract

We explore whether forecasting an aggregate variable using information on its disaggregate components can improve the prediction mean squared error over forecasting the disaggregates and aggregating those forecasts, or using only aggregate information in forecasting the aggregate. An implication of a theory of prediction is that the first should outperform the alternative methods to forecasting the aggregate in population. However, forecast models are based on sample information. The data generation process and the forecast model selected might differ. We show how changes in collinearity between regressors affect the bias-variance trade-off in model selection and how the criterion used to select variables in the forecasting model affects forecast accuracy. We investigate why forecasting the aggregate using information on its disaggregate components improves forecast accuracy of the aggregate forecast of Euro area inflation in some situations, but not in others. The empirical evidence on Euro-zone inflation forecasts suggests that more information can help, more so by including macroeconomic variables than disaggregate components.

JEL: C32, C51, C53, E31

KEYWORDS: Disaggregate information, predictability, forecast model selection, VAR, factor models

---

*Preliminary and incomplete: please do not cite without the authors' permission. We thank seminar participants at Humboldt University and the Free University, Berlin, participants of the ESF-EMM conference, Sardinia, and the CEPR/EABCN workshop on forecasting combination, Brussels, in particular Graham Elliott and Lutz Kilian, for useful comments. Financial support from the ESRC under grant RES 051 270035 is gratefully acknowledged. The views expressed in this paper are not necessarily those of the European Central Bank.

1 Introduction

Forecasts of marcoeconomic aggregates are used by the private sector, governmental and international institutions as well as central banks. Recently there has been renewed interest in the effect of contemporaneous aggregation in forecasting. For example, one issue has been the potential improvement in forecast accuracy delivered by forecasting the component indices and aggregating such forecasts, as against simply forecasting the aggregate itself.[1] The theoretical literature shows that aggregating component forecasts improves over directly forecasting the aggregate if the data generating process is known. If the data generating process is not known and the model has to be estimated, it depends on the unknown data generating process whether the disaggregated approach improves the accuracy of the aggregate forecast. It might be preferable to forecast the aggregate directly. Since in practice the data generating process is not known, it remains an empirical question whether aggregating forecasts of disaggregates improves forecast accuracy of the aggregate of interest. For example, the results in Hubrich (2004) indicate that aggregating forecasts by component does not necessarily help to forecast year-on-year Eurozone inflation twelve months ahead.

In this paper, we suggest an alternative use of disaggregate information to forecast the aggregate variable of interest, that is to include the disaggregate information or disaggregate variables in the model for the aggregate as opposed to forecasting the disaggregate variables separately and aggregating those forecasts.

We show that disaggregating elements of the information set $\mathcal{I}_{T-1}$ into their components cannot lower predictability of a given aggregate $\mathbf{y}_T$. We focus on disaggregation across variables (such as sub-indices of a price measure). Disaggregation may also be considered across space (e.g., regions of an economy), time (higher frequencies), or all of these. The predictability concept considered in this paper concerns a property in population of the variable of interest in relation to an information set. A related predictability concept is discussed by Diebold & Kilian (2001). Whereas that paper considers measuring predictability of different variables based on one information set, we investigate predictability of the same variable based on different information sets. In contrast to predictability as a property in population, we use 'forecastability' to refer to the improvement in forecast accuracy related to the sample information given the unconditional moments of a

[1]See e.g. Espasa, Senra & Albacete (2002), Hubrich (2004) and Benalal, Diaz del Hoyo, Landau, Roma & Skudelny (2004); see also Fair & Shiller (1990) for a related analysis for US GNP). Contributions to the theoretical literature on aggregation versus disaggregation in forecasting can be found in e.g. Grunfeld & Griliches (1960), Kohn (1982), Lütkepohl (1984, 1987), Pesaran, Pierse & Kumar (1989), Van Garderen, Lee & Pesaran (2000); see also Lütkepohl (2004) for a recent review on forecasting aggregated processes by VARMA models.

variable. Potential misspecification of the forecast model due to model selection and estimation uncertainty as well as data measurement errors and structural breaks will affect the accuracy of the resulting forecast and help to explain why theoretical results on predictability are not confirmed in empirical applications (see also Hendry (2004) and Clements & Hendry (2004a)).

In contrast to most previous papers on disaggregation in forecasting that are set up in a VAR framework, our proposal for including disaggregate variables in the aggregate model, gives rise to a classical model selection problem. We analyse model selection and estimation in the conditional model. We condition on We particularly focus on model selection for forecasting, the role of model misspecification in the forecast period and changes in collinearity between regressors. The degree of misspecification, i.e. the deviation of the forecast model from the true data generating process in the forecast period, is not known in practice. Although the predictability theory provides a useful guide for forecasting, we need to empirically investigate the usefulness of different methods of model selection to include disaggregate information for euro area inflation. Thereby we extend the results in Hubrich (2004) and relate our empirical findings to the analytical results presented in the previous sections.

The paper is organised as follows. First, section 2 briefly reviews the notion of (un-)predictability and its properties most relevant to our subsequent analysis. Then we show that adding lagged information on disaggregates to a model of an aggregate must improve predictability. However, an improvement in predictability is a necessary, but not sufficient condition for an improvement in the forecast accuracy. In section 3, we investigate the effect of model selection and estimation uncertainty on the forecast accuracy in a conditional model with particular reference to forecasting the aggregate when disaggregate information is included in the aggregate model. Section 4 notes an extension to dynamic forecasts for horizons larger than one. In section 5, we investigate in a simulated out-of-sample experiment whether adding lagged values of the sub-indices of the Harmonized Index of Consumer Prices (HICP) to a model of the aggregate improves the accuracy of forecasts of that aggregate relative to forecasting the aggregate HICP only using lagged aggregate information, or aggregating forecasts of those sub-indices. Section 6 concludes.

2 Improving predictability by disaggregation

In this section the notion of predictability and its properties most relevant to our subsequent analysis are reviewed first, before we address the issue of predictability and disaggregation.[2]

2.1 Predictability and its properties

A non-degenerate vector random variable ν_t is unpredictable with respect to an information set $\mathcal{I}_{t-1}$ (which always includes the sigma-field generated by the past of ν_t) over a period $\mathcal{T} = \{1, \ldots, T\}$ if its conditional distribution $\mathsf{D}_{\nu_t}(\nu_t|\mathcal{I}_{t-1})$ equals its unconditional $\mathsf{D}_{\nu_t}(\nu_t)$:

$$\mathsf{D}_{\nu_t}(\nu_t \mid \mathcal{I}_{t-1}) = \mathsf{D}_{\nu_t}(\nu_t) \quad \forall t \in \mathcal{T}. \tag{1}$$

Unpredictability, therefore, is a property of ν_t in relation to $\mathcal{I}_{t-1}$ intrinsic to ν_t. Predictability requires combinations with $\mathcal{I}_{t-1}$, as for example in:

$$\mathbf{y}_t = \phi_t(\mathcal{I}_{t-1}, \nu_t) \tag{2}$$

so $\mathbf{y}_t$ depends on both the information set and the innovation component. Then:

$$\mathsf{D}_{\mathbf{y}_t}(\mathbf{y}_t \mid \mathcal{I}_{t-1}) \neq \mathsf{D}_{\mathbf{y}_t}(\mathbf{y}_t) \quad \forall t \in \mathcal{T}. \tag{3}$$

The special case of (2) relevant here (after appropriate data transformations, such as logs) is predictability in mean:

$$\mathbf{y}_t = \mathbf{f}_t(\mathcal{I}_{t-1}) + \nu_t. \tag{4}$$

Other cases of (2) which are potentially relevant are considered in Hendry (2004).

In (4), $\mathbf{y}_t$ is predictable in mean even if ν_t is not as:

$$\mathsf{E}_t[\mathbf{y}_t \mid \mathcal{I}_{t-1}] = \mathbf{f}_t(\mathcal{I}_{t-1}) \neq \mathsf{E}_t[\mathbf{y}_t],$$

in general. Since:

$$\mathsf{V}_t[\mathbf{y}_t \mid \mathcal{I}_{t-1}] < \mathsf{V}_t[\mathbf{y}_t] \quad \text{when} \quad \mathbf{f}_t(\mathcal{I}_{t-1}) \neq \mathbf{0} \tag{5}$$

predictability ensures a variance reduction.

Predictability is obviously relative to the information used. Given an information set, $\mathcal{J}_{t-1} \subset \mathcal{I}_{t-1}$ when the process to be predicted is $\mathbf{y}_t = \mathbf{f}_t(\mathcal{I}_{t-1}) + \nu_t$ as in (4), less accurate predictions will result, but they will remain unbiased. Since $\mathsf{E}_t[\nu_t|\mathcal{I}_{t-1}] = \mathbf{0}$:

$$\mathsf{E}_t[\nu_t \mid \mathcal{J}_{t-1}] = \mathbf{0},$$

[2] The theory of economic forecasting in Clements & Hendry (1998, 1999) for non-stationary processes subject to structural breaks, where the forecasting model differs from the data generating mechanism, is rooted in the properties of (un-)predictability. Hendry (2004) considers the foundations of this predictability concept in more detail.

so that:

$$\mathsf{E}_t\left[\mathbf{y}_t \mid \mathcal{J}_{t-1}\right] = \mathsf{E}_t\left[\mathbf{f}_t\left(\mathcal{I}_{t-1}\right) \mid \mathcal{J}_{t-1}\right] = \mathbf{g}_t\left(\mathcal{J}_{t-1}\right),$$

say. Let $\mathbf{e}_t = \mathbf{y}_t - \mathbf{g}_t\left(\mathcal{J}_{t-1}\right)$, then, providing $\mathcal{J}_{t-1}$ is a proper information set containing the history of the process:

$$\mathsf{E}_t\left[\mathbf{e}_t \mid \mathcal{J}_{t-1}\right] = \mathbf{0},$$

so $\mathbf{e}_t$ is a mean innovation with respect to $\mathcal{J}_{t-1}$.

However, as:

$$\mathbf{e}_t = \left(\mathbf{f}_t\left(\mathcal{I}_{t-1}\right) - \mathbf{g}_t\left(\mathcal{J}_{t-1}\right)\right) + \nu_t = \mathbf{w}_{t-1} + \nu_t$$

(say) where $\mathsf{E}\left[\mathbf{w}_{t-1}\nu_t'\right] = \mathbf{0}$ then:

$$\mathsf{E}_t\left[\mathbf{e}_t \mid \mathcal{I}_{t-1}\right] = \mathbf{f}_t\left(\mathcal{I}_{t-1}\right) - \mathsf{E}_t\left[\mathbf{g}_t\left(\mathcal{J}_{t-1}\right) \mid \mathcal{I}_{t-1}\right] = \mathbf{f}_t\left(\mathcal{I}_{t-1}\right) - \mathbf{g}_t\left(\mathcal{J}_{t-1}\right) \neq \mathbf{0}.$$

As a consequence of this failure of $\mathbf{e}_t$ to be an innovation with respect to $\mathcal{I}_{t-1}$:

$$\begin{aligned}
\mathsf{E}\left[\mathbf{e}_t\mathbf{e}_t'\right] &= \mathsf{E}\left[\left(\nu_t + \mathbf{w}_{t-1}\right)\left(\nu_t + \mathbf{w}_{t-1}\right)'\right] \\
&= \mathsf{E}\left[\nu_t\nu_t'\right] + \mathsf{E}\left[\nu_t\mathbf{w}_{t-1}'\right] + \mathsf{E}\left[\mathbf{w}_{t-1}\nu_t'\right] + \mathsf{E}\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}'\right] \\
&= \mathsf{E}\left[\nu_t\nu_t'\right] + \mathsf{E}\left[\mathbf{w}_{t-1}\mathbf{w}_{t-1}'\right] \\
&\geq \mathsf{E}\left[\nu_t\nu_t'\right]
\end{aligned}$$

where strict equality follows unless $\mathbf{w}_{t-1} = \mathbf{0}\ \forall t$.

Nevertheless, that predictions from $\mathcal{J}_{t-1}$ remain unbiased on the reduced information set suggests that, by itself, incomplete information is not fatal to the forecasting enterprise.

In particular, disaggregating components of $\mathcal{I}_{T-1}$ into their elements cannot lower predictability of a given aggregate $\mathbf{y}_T$, where such disaggregation may be across space (e.g., regions of an economy), time (higher frequency), variables (such as sub-indices of a price measure), or all of these. These attributes suggest forecasting using general models to be a preferable strategy, and provide a formal basis for including as much information as possible, being potentially consistent with many-variable 'factor forecasting' (see e.g. Stock & Watson (2002), and Forni, Hallin, Lippi & Reichlin (2000)), and with the benefits claimed in the 'pooling of forecasts' literature (e.g., Clemen, 1989; Clements & Hendry, 2004b, for a recent theory). Although such results run counter to the common finding in forecasting competitions that 'simple models do best' (see e.g. Makridakis & Hibon, 2000; Allen & Fildes, 2001; Fildes & Ord, 2002), Clements & Hendry (2001) suggest that simplicity is confounded with robustness.

2.2 Predictability and disaggregation

The previous section concerns adding content to the information set $\mathcal{J}_{t-1}$ to deliver $\mathcal{I}_{T-1}$. One form of adding information is via disaggregation of the target variable $\mathbf{y}_T$ into its components $\mathbf{y}_{i,T}$ although $\mathsf{D}_{\mathbf{y}_{T+1}}(\mathbf{y}_{T+1}|\cdot)$ remains the target of interest. We consider only two components and a scalar process to illustrate the analysis, which clearly generalizes to many components and a vector process.

Consider a scalar y_t to be forecast, composed of:

$$y_{T+1} = w_{1,T+1}y_{1,T+1} + w_{2,T+1}y_{2,T+1} \tag{6}$$

with the weights $w_{1,T+1}$ and $w_{2,T+1} = (1 - w_{1,T+1})$ for each of the two components. It may be thought that, when the $y_{i,t}$ themselves depend in different ways on the general information set $\mathcal{I}_{t-1}$, which by construction includes the σ–field generated by the past of the $y_{i,t-j}$, predictability could be improved by forecasting the disaggregates and aggregating those forecasts to obtain those for y_{T+1}. However, let:

$$\mathsf{E}_{T+1}\left[y_{i,T+1} \mid \mathcal{I}_T\right] = \delta'_{i,T+1}\mathcal{I}_T \tag{7}$$

which is the conditional expectation of each component $y_{i,T+1}$ and hence is the minimum mean-square error (MSE) predictor. Then, taking conditional expectations in (6), aggregating the two terms in (7) delivers $\mathsf{E}_{T+1}[y_{T+1}|\mathcal{I}_T]$:

$$\mathsf{E}_T\left[y_{T+1} \mid \mathcal{I}_T\right] = \sum_{i=1}^{2} w_{i,T+1}\mathsf{E}_{T+1}\left[y_{i,T+1} \mid \mathcal{I}_T\right] = \sum_{i=1}^{2} w_{i,T+1}\delta'_{i,T+1}\mathcal{I}_T = \lambda'_{T+1}\mathcal{I}_T \quad \text{(say).}$$

By way of comparison, consider predicting y_{T+1} directly from $\mathcal{I}_T$:

$$\mathsf{E}_{T+1}\left[y_{T+1} \mid \mathcal{I}_T\right] = \phi'_{T+1}\mathcal{I}_T, \tag{8}$$

so $\phi_{T+1} = \lambda_{T+1}$ with a prediction error:

$$y_{T+1} - \mathsf{E}_{T+1}\left[y_{T+1} \mid \mathcal{I}_T\right] = \upsilon_{T+1} \tag{9}$$

which is unpredictable from $\mathcal{I}_T$ and hence nothing is lost predicting y_{T+1} directly instead of aggregating component predictions once the general information set $\mathcal{I}_T$ is used. In practice, if both the weights $w_{i,T+1}$ and the coefficients of the component models $\delta'_{i,T+1}$ change more than the coefficients of the aggregate model λ_{T+1}, forecasting the aggregate directly could well be more accurate than aggregating the component forecasts. Thus, the key issue in (say) aggregate inflation prediction is not predicting the component price changes, but including those components in the information set $\mathcal{I}_T$. This result implies that weights are not needed for aggregating component forecasts, and also saves the additional effort of specifying disaggregate models for the components.

Including the components in the information set $\mathcal{I}_T$ is quite distinct from restricting information to lags of aggregate inflation, an information set we denote by $\mathcal{J}_T$. Then:

$$\mathsf{E}_{T+1}\left[y_{T+1} \mid \mathcal{J}_T\right] = \psi'_{T+1}\mathcal{J}_T,$$

so that using y_{T+1} from (8) and (9) gives:

$$y_{T+1} - \mathsf{E}_{T+1}\left[y_{T+1} \mid \mathcal{J}_T\right] = \phi'_{T+1}\mathcal{I}_T - \psi'_{T+1}\mathcal{J}_T + v_{T+1}, \tag{10}$$

which must have larger MSE than (9), since according to section 2.1, although the predictions based on $\mathcal{I}_T$ and $\mathcal{J}_T$ are both unbiased, the prediction based on the smaller information set $\mathcal{J}_T$, here only including the lags of aggregate inflation and no disaggregate information, is less accurate, and has a larger variance than the forecast based on $\mathcal{I}_T$. If y_{T+1} was unpredictable from both information sets, i.e. $\psi_{T+1} = \phi_{T+1} = \mathbf{0}$, then (9) and (10) would have equal MSE.

2.3 Example

Let the DGP be a vector autoregression of order one in the components $y_{i,t}$:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \tag{11}$$

where $\mathsf{E}[\mathbf{v}_t] = \mathbf{0}$, $\mathsf{E}[\mathbf{v}_t\mathbf{v}'_t] = \mathbf{\Sigma}_v$ and $\mathsf{E}[\mathbf{v}_t\mathbf{v}'_s] = \mathbf{0}$ for all $s \neq t$. Furthermore, $y_t = w_{1,t}y_{1,t} + (1 - w_{1,t})\, y_{2,t}$, as in a price index, where weights shift with value shares, leading to:

$$\begin{aligned} y_t \quad = \quad & w_{1,t}\left[(\pi_{11} - \pi_{21})\, y_{1,t-1} + (\pi_{12} - \pi_{22})\, y_{2,t-1}\right] + \pi_{21}y_{1,t-1} + \pi_{22}y_{2,t-1} \\ & + w_{1,t}v_{1,t} + (1 - w_{1,t})\, v_{2,t}. \end{aligned} \tag{12}$$

2.3.1 Disaggregate forecasting model: True disaggregate process known

The disaggregate forecasting model for known parameters is:

$$\begin{pmatrix} \widehat{y}_{1,T+1|T} \\ \widehat{y}_{2,T+1|T} \end{pmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} \begin{pmatrix} y_{1,T} \\ y_{2,T} \end{pmatrix},$$

with:

$$\widehat{y}_{T+1|T} = w_{1,T+1}\widehat{y}_{1,T+1|T} + w_{2,T+1}\widehat{y}_{2,T+1|T}.$$

Thus, the forecast error from forecasting the disaggregate components and aggregating those forecasts is:

$$\begin{aligned} y_{T+1} - \widehat{y}_{T+1|T} &= w_{1,T+1}\left(y_{1,T+1} - \widehat{y}_{1,T+1|T}\right) + w_{2,T+1}\left(y_{2,T+1} - \widehat{y}_{2,T+1|T}\right) \\ &= w_{1,T+1}v_{1,T+1} + w_{2,T+1}v_{2,T+1} \end{aligned} \tag{13}$$

which is unpredictable, independent of whether the weights are known or not known.

2.3.2 Aggregate forecasting model with known parameters: True disaggregate process known

In contrast to the first example where the disaggregate forecasting model is fitted to the process, consider restricting the information set underlying the forecasting model to lags of y_t alone, with no disaggregates used. Furthermore, the true aggregate process is assumed known so that the true parameters of the aggregate forecasting model are known to the forecaster. In the following, to simplify the presentation, it is assumed that $w_{1,t} = w_{2,t} = 1$,[3] so that $y_t = y_{1,t} + y_{2,t}$. Then the aggregate y_t based on the true disaggregate process (11) can be represented by an ARMA(2,1) process (for a proof see e.g. Lütkepohl, 1987, Ch.4,1984a).

The VAR in (11) can be written as $\mathbf{\Pi}(L)\mathbf{y}_t = \mathbf{v}_t$:

$$\begin{pmatrix} 1-\pi_{11}L & -\pi_{12}L \\ -\pi_{21}L & 1-\pi_{22}L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}. \tag{14}$$

Multiplying (14) by the adjoint $\mathbf{\Pi}(L)^*$ of the VAR operator $\mathbf{\Pi}(L)$ gives:

$$\begin{pmatrix} 1-a_1-a_2L^2 & 0 \\ 0 & 1-a_1-a_2L^2 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1-\pi_{22}L & \pi_{12}L \\ \pi_{21}L & 1-\pi_{11}L \end{pmatrix} \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}. \tag{15}$$

Furthermore, multiplying (15) by the vector of weights $F = (1, 1)$ of the disaggregate components entails:

$$(1 - a_1L - a_2L^2)y_t = (1 - b_1L)v_{1,t} + (1 - b_2L)v_{2,t} \tag{16}$$

with $a_1 = \pi_{11} + \pi_{22}$, $a_2 = \pi_{12}\pi_{21} - \pi_{11}\pi_{22}$, $b_1 = \pi_{21} + \pi_{22}$ and $b_2 = \pi_{12} + \pi_{11}$. It can be shown that the right-hand side of expression (16) is a process with an MA(1) representation, so that the aggregate process has an ARMA(2,1) representation: $(1 - a_1L - a_2L^2)y_t = (1 - \gamma L)u_t$.[4]

[3] Results are easily extended to the case of different and time-varying component weigths.

[4] More generally, it has been shown in the literature that, if the disaggregate process follows a VARMA(p, q), the aggregate process follows an ARMA(p^*, q^*) process with $p^* \leq (n-m) + 1 \times p$ and $q^* \leq (n-m) \times p + q$ with n being the number of variables in the system and m being the rank of the matrix of aggregation weights (see e.g. Lütkepohl, 1987, Ch.4).

The model in (16) is used as a forecasting model based on the information set restricted to the aggregate:

$$\hat{y}_{T+1} = a_1 y_T + a_2 y_{T-1} + v_{1,T+1} - b_1 v_{1,T} + v_{2,T+1} - b_2 v_{2,T} \tag{17}$$

To derive the forecast error made, recall that the aggregate is $y_t = y_{1,t} + y_{2,t}$. Then (16) entails:

$$\begin{aligned} y_t &= a_1 y_{1,t-1} + a_2 y_{1,t-2} + a_1 y_{2,t-1} + a_2 y_{2,t-2} \\ &\quad + v_{1,t} - b_1 v_{1,t-1} + v_{2,t} - b_2 v_{2,t-1} \end{aligned} \tag{18}$$

Since in this section, we have assumed that $w_{1,t} = w_{2,t} = 1$ for ease of exposition, the disaggregate process in (11) simplifies to

$$y_t = (\pi_{11} + \pi_{21})\, y_{1,t-1} + (\pi_{12} + \pi_{22})\, y_{2,t-1} + v_{1,t} + v_{2,t}. \tag{19}$$

Then the forecast error of the disaggregate process is given by the difference between (19) and (18):

$$\begin{aligned} \widehat{u}_{T+1|T} &= y_{T+1} - \widehat{y}_{T+1|T} \\ &= (\pi_{11} + \pi_{21})\, y_{1,T} - a_1 y_{1,T} - a_2 y_{1,T-1} \\ &\quad + (\pi_{12} + \pi_{22})\, y_{2,T} - a_1 y_{2,T} - a_2 y_{2,T-1} \\ &\quad + v_{1,T+1} + v_{2,T+1} - v_{1,T+1} + b_1 v_{1,T} - v_{2,T+1} + b_2 v_{2,T} \\ &= (\pi_{11} - \pi_{22})\, y_{1,T} - a_2 y_{1,T-1} + (\pi_{12} - \pi_{11})\, y_{2,T} - a_2 y_{2,T-1} \\ &\quad + b_1 v_{1,T} + b_2 v_{2,T} \end{aligned}$$

which will not be unpredictable in general. The entailed restrictions are of the following form[5]:

$$\begin{aligned} \pi_{21} - \pi_{22} &= 0 \\ \pi_{12} - \pi_{11} &= 0 \\ a_2 &= -\pi_{11}\pi_{22} - \pi_{12}\pi_{21} = 0 \end{aligned}$$

These restrictions will usually not be fulfilled simultaneously, so u_t will be predictable from $y_{1,t-i}$ and/or $y_{2,t-i}$ $(i = 1, 2)$.

2.3.3 Aggregate forecasting model with unknown parameters: True disaggregate process not known

Alternatively, consider again restricting the information set to lags of y_t with no disaggregates used. In contrast to the previous example, the true disaggregate process is not known. Consequently, the aggregate process has to be approximated. A

[5](See e.g., Lütkepohl, 1984, for the implied restrictions for equality of the aggregate and the disaggregate forecast model for a more general DGP).

further difference to the previous example is that we assume that the aggregate is a weighted average of the two disaggregates where the weights are allowed to vary across components and change over time.

We approximate (11) by an autoregression of the form:

$$y_t = \rho y_{t-1} + u_t \tag{20}$$

where:

$$\widehat{y}_{T+1|T} = \widehat{\rho} y_T.$$

Since $y_t = w_{1,t} y_{1,t} + (1 - w_{1,t}) y_{2,t}$, (20) entails that:

$$y_t = \rho w_{1,t-1} y_{1,t-1} + \rho (1 - w_{1,t-1}) y_{2,t-1} + u_t. \tag{21}$$

Thus, the forecast error $\widehat{u}_{T+1|T}$ from forecasting the true disaggregate process (11) with an estimated AR(1) model is given by (12) minus (21):

$$\begin{aligned} \widehat{u}_{T+1|T} &= y_{T+1} - \widehat{y}_{T+1|T} \\ &= (w_{1,T+1} [\pi_{11} - \pi_{21}] + \pi_{21} - \widehat{\rho} w_{1,T}) y_{1,T} \\ &\quad + (w_{1,T+1} [\pi_{12} - \pi_{22}] + \pi_{22} - \widehat{\rho} (1 - w_{1,T})) y_{2,T} \\ &\quad + w_{1,T+1} v_{1,T+1} + (1 - w_{1,T+1}) v_{2,T+1}, \end{aligned} \tag{22}$$

which will not be unpredictable in general. Even for constant weights, the entailed restrictions are well known to be of the form:

$$\begin{aligned} w_1 (\pi_{11} - \pi_{21} - \widehat{\rho}) + \pi_{21} &= 0 \\ w_1 (\pi_{12} - \pi_{22} + \widehat{\rho}) + (\pi_{22} - \widehat{\rho}) &= 0 \end{aligned}$$

There is no reason to anticipate that $\widehat{\rho}$ can simultaneously satisfy both requirements (even less so with time-varying weights), so u_{T+1} will be predictable from $y_{1,T}$ and $y_{2,T}$, as in the previous example where the true aggregate process was known.

These results indicate that it should improve forecast accuracy to include disaggregate information in the aggregate forecasting model. The additional difficulties in an actual forecast exercise of the choice of the information set, estimation of unknown parameters, unmodeled breaks, forecasting the weights, and data measurement errors that the forecaster faces, however, may be sufficiently large to offset the potential benefits. The role of estimation and model selection in a conditional model is considered analytically in the next section and is extended to a very simple dynamic forecasts in section 4.1. Section 5 presents an empirical analysis for forecasting euro-area inflation.

3 Model selection and estimation in a conditional model

When forecasting aggregates by disaggregates, the selection issue concerns retaining or omitting the disaggregates. First, a general framework is considered in this section to establish the role of model selection and estimation in a conditional model. Then we relate the discussion to the choice of a forecasting model for forecasting aggregates by disaggregates.

We first state our notation. Consider the conditional regression model:

$$y_t = \beta' \mathbf{x}_t + \nu_t \text{ where } \nu_t \sim \mathsf{IN}\left[0, \sigma_\nu^2\right] \tag{23}$$

with $\mathbf{x}_t \sim \mathsf{IN}_k\left[\mu, \mathbf{\Sigma}\right]$ (independent normal, mean μ, variance $\mathbf{\Sigma}$) independently of $\{\nu_t\}$. Then using $\mathsf{E}\left[\cdot\right]$ to denote an expectation:

$$\mathsf{E}\left[\mathbf{x}_t \mathbf{x}_t'\right] = \mu\mu' + \mathbf{\Sigma} = \mathbf{\Omega}, \tag{24}$$

say with $\mathsf{E}\left[y_t\right] = \beta'\mu$. The notation allows that one of the components of $\mathbf{x}_t$ is a unit vector. For large T, it is well known (see e.g., Hendry (1995) that $\mathsf{E}[\widehat{\beta}] = \beta$ and:

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \underset{a}{\sim} \mathsf{N}_k\left[\mathbf{0}, \sigma_\nu^2 \mathbf{\Omega}^{-1}\right], \tag{25}$$

so ($\mathsf{V}\left[\cdot\right]$ denotes a variance):

$$\mathsf{V}\left[\widehat{\beta}\right] \simeq \sigma_v^2 T^{-1} \mathbf{\Omega}^{-1}.$$

For k regressors with estimated coefficients and a known outside-sample value of $\mathbf{x}$, denoted $\mathbf{x}_{T+1}$, a forecast can be based on:

$$\widehat{y}_{T+1} = \widehat{\beta}' \mathbf{x}_{T+1}, \tag{26}$$

when:

$$y_{T+1} = \beta' \mathbf{x}_{T+1} + \nu_{T+1}, \tag{27}$$

with a forecast error:

$$\widehat{\nu}_{T+1} = y_{T+1} - \widehat{y}_{T+1} = \left(\beta - \widehat{\beta}\right)' \mathbf{x}_{T+1} + \nu_{T+1} \tag{28}$$

so that its conditional mean-square forecast error (MSFE) is (letting $\mathsf{E}\left[\nu_{T+1}\right] = 0$ and $\mathsf{E}\left[\nu_{T+1}^2\right] = \sigma_\nu^2$):

$$\mathsf{M}\left[\widehat{\nu}_{T+1} \mid \mathbf{x}_{T+1}\right] = \sigma_v^2 + \mathbf{x}_{T+1}' \mathsf{V}\left[\widehat{\beta}\right] \mathbf{x}_{T+1} \simeq \sigma_\nu^2 \left(1 + T^{-1} \mathbf{x}_{T+1}' \mathbf{\Omega}^{-1} \mathbf{x}_{T+1}\right). \tag{29}$$

The unconditional MSFE is obtained by taking the expectation of (29) over drawings of $\mathbf{x}_{T+1}$ from its distribution, mainly using (24), in which case, (29) simplifies to the well-known formula:

$$\mathsf{M}\left[\widehat{\nu}_{T+1}\right] = \sigma_{\nu}^{2}\left(1 + T^{-1}k\right). \tag{30}$$

Even in a constant parameter setting, to judge selection by (29) requires several aspects that do not seem to have been addressed, and are discussed below. First, what criteria should be used to judge the parsimony of the selected model? Researchers often use 'statistical significance', determined by the conventional rule that an observed t-statistic exceeds its 5% significance level, which is approximately 2. To avoid having to consider signs, we translate that criterion into $\mathsf{t}^2_{\widehat{\beta}_i} \geq 4$ for the *i*-th variable. Since collinearity is likely to influence the observed t-value, we first analyze that issue, then consider the converse problem of omitting relevant variables, and finally combine these analyses to try to determine rules for model selection.

The assumptions underlying the conditional MSFE (29) are quite strong since they imply parameter constancy. For an extensive analysis of such problems of different parameter values in the forecast regime the reader is referred to Clements & Hendry (1999).

3.1 Collinearity

Factorize the variance-covariance matrix $\mathbf{\Omega}$ of the regressors x_t in (25) as $\mathbf{\Omega} = \mathbf{H}'\mathbf{\Lambda}\mathbf{H}$ where $\mathbf{H}'\mathbf{H} = \mathbf{I}_k$ and $\mathbf{H}\mathbf{x}_t = \mathbf{z}_t$, so that:

$$\mathsf{E}\left[\mathbf{z}_t\mathbf{z}_t'\right] = \mathbf{\Lambda} \tag{31}$$

and:

$$y_t = \gamma'\mathbf{z}_t + \nu_t \text{ where } \gamma = \mathbf{H}\beta. \tag{32}$$

Clearly, despite the transform, $\widehat{\gamma} \equiv \mathbf{H}\widehat{\beta}$ and:

$$\widehat{y}_{T+1} = \widehat{\gamma}'\mathbf{z}_{T+1} = \widehat{\beta}'\mathbf{H}'\mathbf{H}\mathbf{x}_{T+1} = \widehat{\beta}'\mathbf{x}_{T+1},$$

so neither estimation nor forecasting are affected. Then:

$$\begin{aligned}
\mathbf{x}_{T+1}'\mathbf{\Omega}^{-1}\mathbf{x}_{T+1} &= \mathbf{x}_{T+1}'\left(\mathbf{H}'\mathbf{\Lambda}\mathbf{H}\right)^{-1}\mathbf{x}_{T+1} = \mathbf{x}_{T+1}'\mathbf{H}'\mathbf{\Lambda}^{-1}\mathbf{H}\mathbf{x}_{T+1} \\
&= \mathbf{z}_{T+1}'\mathbf{\Lambda}^{-1}\mathbf{z}_{T+1} = \sum_{i=1}^{k}\frac{z_{i,T+1}^{2}}{\lambda_i}.
\end{aligned} \tag{33}$$

On average, (31) entails that $\mathsf{E}[z_{i,T+1}^2] = \lambda_i$, and therefore, unconditionally:

$$\mathsf{E}\left[\mathbf{x}_{T+1}'\mathbf{\Omega}^{-1}\mathbf{x}_{T+1}\right] = \mathsf{E}\left[\sum_{i=1}^{k}\frac{z_{i,T+1}^2}{\lambda_i}\right] = k. \tag{34}$$

Thus, substituting (34) in the unconditional value of (29) again simplifies to (30). That result shows that any 'collinearity' in $\mathbf{x}_t$ is irrelevant to forecasting, *so long as the marginal process remains constant.* Alternatively, the linear regression model is invariant under linear, and therefore orthogonal transforms, as shown in (32), so collinearity is not an attribute of a model but only of a particular parameterization of that model.

3.1.1 Changes in collinearity in the forecast period

When β stays constant, but the regressor variance-covariance matrix $\mathbf{\Omega}$ of the in-sample period changes to $\mathbf{\Omega}^*$ out-of-sample, so the mean square of the marginal process $\mathbf{x}_{T+1}$ alters, with $\mathbf{\Lambda}$ changing to $\mathbf{\Lambda}^*$ then:

$$\mathsf{E}\left[\sum_{i=1}^{k}\frac{z_{i,T+1}^2}{\lambda_i}\right] = \sum_{i=1}^{k}\frac{\lambda_i^*}{\lambda_i}, \tag{35}$$

so that the unconditional MSFE is:

$$\mathsf{M}\left[\widehat{\nu}_{T+1}\right] \simeq \sigma_\nu^2\left(1 + T^{-1}\sum_{i=1}^{k}\frac{\lambda_i^*}{\lambda_i}\right). \tag{36}$$

Changes in the magnitude of the eigenvalue of the least well determined β_j, corresponding to the smallest λ_j, will induce the biggest relative changes in $\mathsf{M}[\widehat{\nu}_{T+1}]$. For example, let the smallest $\lambda_j = 0.0001$ where the change is to $\lambda_j^* = 0.01$ which remains small. Nevertheless, $\lambda_j^*/\lambda_j = 100$ rather than unity, so a dramatic increase in the MSFE arises from retaining that variable.

3.2 Mis-specification

Consider a model based on prior simplification which happens to exclude a regressor set $\mathbf{x}_{2,t}$, where we partition $\mathbf{x}_t' = (\mathbf{x}_{1,t}' : \mathbf{x}_{2,t}')$ of dimensions k_1 and k_2 respectively, when $k_1 + k_2 = k$, leading to the forecast:

$$\overline{y}_{T+1} = \mathbf{x}_{1,T+1}'\overline{\beta}_1, \tag{37}$$

where $\beta' = (\beta_1' : \beta_2')$ and:

$$\overline{\beta}_1 = \left(\sum_{t=1}^{T} \mathbf{x}_{1,t} \mathbf{x}_{1,t}' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_{1,t} y_t \right). \tag{38}$$

Without loss for an analysis of forecasting, we consider the case where $\mathbf{\Omega} = \mathbf{\Lambda}$, so that:[6]

$$\beta_{1,e} = \mathsf{E}[\overline{\beta}_1] \simeq \mathbf{\Lambda}_{11}^{-1} \rho_{1y} = \beta_1 + \mathbf{\Lambda}_{11}^{-1} \mathbf{\Lambda}_{12} \beta_2 = \beta_1, \tag{39}$$

where $\rho_{1y} = \mathsf{E}[\mathbf{x}_{1,t} y_t]$. The forecast error resulting at $T+1$ is $\overline{\omega}_{T+1|T} = y_{T+1} - \overline{y}_{T+1|T}$ so:

$$\overline{\omega}_{T+1|T} = \mathbf{x}_{1,T+1}' \left(\beta_1 - \overline{\beta}_1 \right) + \mathbf{x}_{2,T+1}' \beta_2 + \omega_{T+1}, \tag{40}$$

with expectation:

$$\mathsf{E} \left[\overline{\omega}_{T+1|T} \mid \mathbf{x}_{T+1} \right] = \mathbf{x}_{1,T+1}' \left(\beta_1 - \beta_{1,e} \right) + \mathbf{x}_{2,T+1}' \beta_2 = \mathbf{x}_{2,T+1}' \beta_2. \tag{41}$$

The forecast is systematically biased by $\mathbf{x}_{2,T+1}' \beta_2$. When $\mathbf{x}_{2,T+1}' \beta_2 \simeq \mathbf{x}_{2,t}' \beta_2$ $\forall t = 1, \ldots, T$, this bias is 'absorbed by', and reflected in, the in-sample estimates, so the forecast MSFE is close to that anticipated from the in-sample estimates.. However, if $\mathbf{x}_{2,T+1}'$ differs markedly from in-sample values, then serious forecast errors could result. Therefore, in a non-constant parameter process a change even in the excluded variables matters for forecasting.

3.3 Mis-specification and collinearity

In this section, we derive the trade-off between the estimation costs from retaining a variable–leading to an increased forecast variance–and the increased bias due to mis-specification costs when a variable is incorrectly omitted. We consider a process with changing collinearity: note that under general linear transformations, that would become a non-constant parameter process.

The conditional variance of the forecast error $\overline{\omega}_{T+1|T}$ due to the omission of a relevant variable is:

$$\mathsf{E} \left[\left(\overline{\omega}_{T+1|T} - \mathsf{E} \left[\overline{\omega}_{T+1|T} \mid \mathbf{x}_{T+1} \right] \right)^2 \mid \mathbf{x}_{T+1} \right] = \sigma_\omega^2, \tag{42}$$

[6] A similar result holds under non-orthogonal regressors on replacing $\mathbf{\Lambda}_{22}$ below by $\mathbf{\Omega}_{22} - \mathbf{\Omega}_{21} \mathbf{\Omega}_{11}^{-1} \mathbf{\Omega}_{12}$, and retaining covariances.

so using the factorizations from section 3.1, the conditional MSFE is (tr denotes the trace):

$$\begin{aligned}\mathsf{E}\left[\overline{\omega}^2_{T+1|T} \mid \mathbf{x}_{T+1}\right] &= \sigma^2_\omega + tr\left\{\mathsf{E}\left[\mathbf{x}_{1,T+1}\mathbf{x}'_{1,T+1}\right]\mathsf{V}\left[\overline{\beta}_1\right]\right\} + \beta'_2\mathsf{E}\left[\mathbf{x}_{2,T+1}\mathbf{x}'_{2,T+1}\right]\beta_2 \\ &= \sigma^2_\omega\left(1 + T^{-1}\sum_{i=1}^{k_1}\frac{\lambda^*_i}{\lambda_i}\right) + \beta'_2\mathbf{\Lambda}^*_{22}\beta_2. \qquad (43)\end{aligned}$$

Thus, the net costs of inappropriate exclusion are given by the last term relative to the 'saving' from not having the estimation cost of:

$$T^{-1}\sum_{i=k_1+1}^{k}\frac{\lambda^*_i}{\lambda_i}.$$

This trade-off is central to selecting a particular specification from a class when the objective is forecasting, albeit that the future values of the λ^*_i are bound to be unknown, and is explored in more detail in section 4.

3.4 Adding disaggregates to forecast aggregates

Let $\mathbf{y}_t$ denote the vector of n disaggregate prices with elements $y_{i,t}$ where we illustrate using:

$$\mathbf{y}_t = \mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{e}_t \qquad (44)$$

as the DGP for the disaggregates. Let $y_t = \omega'_t\mathbf{y}_t$ be the aggregate price index with weights ω_t. Then pre-multiplying (44) by ω'_t:

$$\begin{aligned} y_t &= \omega'_t\mathbf{\Gamma}\mathbf{y}_{t-1} + \omega'_t\mathbf{e}_t = \kappa\omega'_{t-1}\mathbf{y}_{t-1} + \left(\omega'_t\mathbf{\Gamma} - \kappa\omega'_{t-1}\right)\mathbf{y}_{t-1} + \omega'_t\mathbf{e}_t \\ &= \kappa y_{t-1} + \left(\phi_t - \kappa\omega_{t-1}\right)'\mathbf{y}_{t-1} + \nu_t \qquad (45)\end{aligned}$$

where $\phi_t = \mathbf{\Gamma}\omega_t$. Thus, even if the DGP is (44) at the level of the components, an aggregate model will be systematically improved by adding disaggregates only to the extent that $\phi_t - \kappa\omega_{t-1} = \pi$ is both constant, and elements contribute substantively to the explanation.

To relate (45) to (23), let $z_t = \rho'\mathbf{x}_t$ be an aggregate variable (such as a 'factor' component), so the analysis applies more widely than just price index aggregation, then (23) can be expressed as:[7]

$$y_t = \kappa z_t + \left(\beta - \kappa\rho\right)'\mathbf{x}_t + u_t. \qquad (46)$$

[7] Even though (46) is perfectly collinear if the price index weights are unchanging, either any component can be deleted or a *PcGets* model selection approach can be adopted (see Hendry & Krolzig, 2004) .

This requires that β is constant, and the choice of aggregate is a constant linear function of $\mathbf{x}_t$. From (45), those conditions seem unlikely to be fulfilled when the regressors $\mathbf{x}_t$ in (46) are $\mathbf{y}_{t-1}$ and z_t is y_{t-1} with $\rho = \omega_{t-1}$. If so, the additional role of disaggregate information over just the aggregate in (46) is represented by the extent to which $\beta \neq \kappa\rho$ for each variable. This is similar to (45), with the added requirement of constancy of π.

Four distinct types of change can be distinguished in (45):

a) changes in the price index weights ω_{t-1} can be due to changes in expenditure shares with constant correlations between the disaggregates;

b) changes in the second-moment matrix of the disaggregates $\mathbf{y}_{t-1}$ (i.e., in the regressor correlation structure) can change collinearity, inducing the effects noted in the previous section;

c) changes in the parameters ϕ_t of the disaggregates, so the role of the disaggregate regressors is non-constant; and

d) changes in the autoregressive parameter κ.

All four potential shifts influence the decision of whether or not to include (or model) the disaggregates, and might hamper possible improvements in the forecast of the aggregate y_t from adding disaggregate variables $x_{i,t}$ to a model with lags of the aggregate already included. The first three of these shifts favour an aggregate model, and could do so even if κ is not constant. The selection issue in this context, therefore, concerns omitting or retaining the disaggregates, where the trade-off is between the impact of changing collinearity increasing forecast uncertainty as described in section 3.1.1, and the mis-specification costs of omitting relevant regressors considered in section 3.2. We evaluate that trade-off in a non-constant-parameter process in section 4, which establishes that for retention of disaggregate variables to be useful, the non-centralities of their squared t-statistics in (45) must be greater than unity.

Given these results, we now consider the selection of a forecasting model.

4 Model selection for forecasting

The central issue of this paper is how to select a forecasting model given the considerations discussed above. When forecasting aggregates by disaggregates, the selection issue concerns retaining or omitting the disaggregates. First, a general framework is considered in this section to establish the role of model selection and estimation in a conditional model. Then we relate the discussion to the choice of a forecasting model for forecasting aggregates by disaggregates.

If all relevant variables were known and only those were included, and the observation to be forecast was a random draw from the same population as the estimation sample, then (29) would correctly represent the resulting MSFE. If the 'most-accurate forecast' is defined by minimizing $\mathsf{M}\left[\widehat{\nu}_{T+1}|\mathbf{x}_{T+1}\right]$, then criteria for trying to select that model can be determined. The central trade-off is between retaining or omitting variables that will improve the accuracy of the forecast mean, where retention will add to the forecast-error variance. Notice the deliberate use of the word 'will': even under the conditions stated here, (26) depends on the actual forecast origin (i.e., $\mathbf{x}_{T+1}$). When $\mathbf{x}_{T+1}$ can differ markedly from any in-sample observed value, model choice can differ from that conventionally supposed, and that issue is addressed below.

To relate the MSFE conditional on x_{T+1} in (43) to the unconditional MSFE in (36), we express the former in terms of the non-centrality $\tau^2_{\beta_j}$ of the expected values of the t^2-tests on the $\beta_{2,i}$ in the DGP noting that:

$$\beta_2'\mathbf{\Lambda}_{22}\beta_2 = \sum_{i=k_1+1}^{k} \beta_i^2 \lambda_i. \tag{47}$$

Let the non-centrality $\tau^2_{\beta_j}$ of the t^2-tests on β_j be expressed in terms of the true parameters of the DGP:

$$\tau^2_{\beta_j} = \frac{\beta_j^2 \sum_{t=1}^{T} x^2_{j,t}}{\sigma^2_\omega} \simeq \frac{T\beta_j^2\lambda_j}{\sigma^2_\omega}. \tag{48}$$

From (43) and (48), the conditional MSFE $\mathsf{E}[\overline{\omega}^2_{T+1|T}|\mathbf{x}_{T+1}]$ from using only the first set of regressors $\mathbf{x}_{1,T+1}$ and omitting the second $\mathbf{x}_{2,T+1}$ can be related to the conditional MSFE $\mathsf{E}[\widehat{\omega}^2_{T+1|T}|\mathbf{x}_{T+1}]$ from using all the variables as:

$$\begin{aligned}
\mathsf{E}\left[\overline{\omega}^2_{T+1|T} \mid \mathbf{x}_{T+1}\right] &\simeq \sigma^2_\omega \left(1 + T^{-1}\sum_{i=1}^{k_1}\frac{\lambda_i^*}{\lambda_i} + T^{-1}\sum_{j=k_1+1}^{k}\frac{T\beta_j^2\lambda_j^*}{\sigma^2_\omega}\right) \\
&= \sigma^2_\omega \left(1 + T^{-1}\left[\sum_{i=1}^{k_1}\frac{\lambda_i^*}{\lambda_i} + \sum_{j=k_1+1}^{k}\tau^2_{\beta_j}\frac{\lambda_j^*}{\lambda_j}\right]\right) \\
&= \mathsf{E}\left[\widehat{\omega}^2_{T+1|T} \mid \mathbf{x}_{T+1}\right] + \frac{\sigma^2_\omega}{T}\sum_{j=k_1+1}^{k}\left(\tau^2_{\beta_j} - 1\right)\frac{\lambda_j^*}{\lambda_j}.
\end{aligned} \tag{49}$$

For a simplified model to out-perform relative to the unrestricted requires that the average $\tau^2_{\beta_j}$ be less than unity for the omitted variables. In such a case, a selected model can out-perform the estimated DGP, as well as less parsimonious estimated models. Omitting irrelevant variables ($\tau^2_{\beta_j} = 0$) is clearly sufficient to justify selecting the simplified model, but is not necessary. If there is no change in collinearity between the regressors in the forecast period in comparison with the in-sample period, i.e., $\lambda^*_j = \lambda_j$, the cost-benefit trade-off from (43) relative to (36) only depends on $\sum_{j=k_1+1}^{k} \tau^2_{\beta_j}$ relative to k_2, but (49) reveals considerably more structure to the choice. Indeed, when collinearity between the regressors is changing in the forecast period, e.g., $\lambda^*_j \neq \lambda_j$, then the general formula in (49) is needed. Changes in collinearity affect forecasts from both included and incorrectly-excluded variables, so simply omitting highly collinear regressors is not a viable solution for selecting forecasting models.

For the illustrative case of $k_2 = 1$ (which implies omitting one relevant regressor) then the cost (or benefit) of omission is simply:

$$T^{-1}\sigma^2_\omega \left(\tau^2_{\beta_k} - 1\right) \frac{\lambda^*_k}{\lambda_k}. \tag{50}$$

If that variable is irrelevant, so $\tau^2_{\beta_k} = 0$, and $\lambda^*_k > \lambda_k$ by a ratio of 100 (say), there is a dramatic benefit to correct omission. Conversely, if $\tau^2_{\beta_k} = 2$ (as the symmetric case), an equally large loss occurs. However, the probability of retaining such a variable on t-testing also rises with $\tau^2_{\beta_k}$.

Notice that (49) is based on the $\tau^2_{\beta_k}$ in the DGP, not the observed $\mathsf{t}^2_{\beta_k=0}$ that happened to arise in-sample when estimating the model. Because the $\mathsf{t}^2_{\beta_k=0}$ have a sampling distribution even in the DGP, selecting those variables that happen to have observed $\mathsf{t}^2_{\beta_k=0}$ values larger than unity will lower *calculated* MSFE (as shown above), but will not lower the true MSFE unless they correspond to $\tau^2_{\beta_j} > 1$. Of course, $\tau^2_{\beta_k}$ is not observed, only $\mathsf{t}^2_{\beta_k=0}$. Under the null that $\beta_k = 0$, the squared t-test, $\mathsf{t}^2_{\beta_k=0}$ is:

$$\mathsf{t}^2_{\beta_k=0} = \frac{T\widehat{\beta}^2_k \sum_{t=1}^{T} x^2_{k,t}}{\widehat{\sigma}^2_\omega} \sim \chi^2_1(0), \tag{51}$$

so:

$$\mathsf{E}\left[\mathsf{t}^2_{\beta_k=0} \mid \tau^2_{\beta_k} = 0\right] = \mathsf{E}\left[\chi^2_1(0)\right] = 1.$$

Consequently (see () Johnson & Kotz, 1970):

$$\mathsf{E}\left[\mathsf{t}^2_{\beta_k=0} \mid \tau^2_{\beta_k} \neq 0\right] \simeq 1 + \tau^2_{\beta_k}. \tag{52}$$

Thus, although retention under the null would be 50% for a criterion where $\mathsf{t}^2_{\beta_k=0} > 1$ (only one tail of the distribution is relevant since $\beta_k \neq 0$ can have either a positive or a negative sign), the result in (52) suggests requiring $\mathsf{t}^2_{\beta_k=0} > 2$ as correct cut-off for retaining the associated variable.

In the present context, the disaggregate prices are inter-correlated, and seem to have changing collinearity, with weights that are also changing in the price index. In the empirical study, the weights in first year of the out-of-sample period change little, but change between 0% for processed food inflation up to almost 6.5% for energy inflation in 1999 over the previous (last in-sample period for 12 months ahead forecasts) year. Indeed, collinearity between components is changing from the in-sample to the out-of-sample period (and also over the forecast period).

Should some of the disaggregates have non-zero effects in (46), the costs of omitting them are the mis-specification effects in (49). We assume their net effects, after including the aggregate information, is not too large, say $\tau^2_{\beta_k} \leq 9$. Retaining a variable with a value of $\tau^2_{\beta_k} = 1$ carries the same cost as omitting it, namely $T^{-1}\sigma^2_{\omega}\lambda^*_k/\lambda_k$: smaller $\tau^2_{\beta_k}$ favour omission, larger favour retention. If the DGP were by chance specified as the initial model, but tested in a conventional manner, then variables with $\tau^2_{\beta_{k=0}} = 2$ would be retained on average (i.e., half the time) only if $\alpha = 0.16$. Interestingly, that significance value is close to the implicit significance level of AIC when $T = 100$ for a range of N (see Akaike (1973) and Campos, Hendry & Krolzig (2003)).

The probability-weighted data-based omission costs from (49) when $\tau^2_{\beta_k} = r$ are:

$$\mathsf{p}\left(\mathsf{t}^2_{\beta_{k(DGP)}} < c_\alpha \mid \tau^2_{\beta_k} = r; \alpha\right) T^{-1}\sigma^2_{\omega}\left(r-1\right)\frac{\lambda^*_k}{\lambda_k}.$$

Table 1 reports these costs for integer values of $2 \leq \tau^2_{\beta_k} \leq 9$, with $\alpha = 0.16$, and reveals that they are remarkably constant across the range of relevant $\tau^2_{\beta_k}$ values considered.

Table 1: Probability-weighted omission costs

$\tau^2_{\beta_k} = r$	$(r-1) \times \mathsf{p}\left(\mathsf{t}^2_{\beta_{k(DGP)}} < c_\alpha \mid \tau^2_{\beta_k} = r; \alpha = 0.16\right)$	cost
2	1×0.50	0.50
3	2×0.37	0.74
4	3×0.28	0.84
5	4×0.21	0.84
6	5×0.15	0.75
7	6×0.11	0.66
8	7×0.08	0.56
9	8×0.06	0.48

Conversely, because a large number of candidate regressors is created by entering all the disaggregates at several lags, adventitious significance is likely: with N irrelevant candidates and a significance level of α, then αN will be retained by chance. If such a loose significance level as $\alpha = 0.16$ were used for a general model with $N = 20$, then 3 irrelevant variables would be retained on average, with costs of $\sum_{k=1}^{3} \sigma_\omega^2 T^{-1} \lambda_k^* / \lambda_k$. This corresponds roughly to the total effect for four terms from table 1. A larger value of α than 0.16 would lower the costs in table 1 and raise the retention rate for irrelevant variables, suggesting perhaps seeking equality for a given N to minimize their total costs.

4.1 Dynamic forecasts

The analysis of open models, where some variables are not endogenized, is difficult primarily because the properties of the 'explanatory' variables are unspecified. In practice, for multi-step ahead forecasts, either those variables have to be forecasted 'off-line', which is reliant on untested strong exogeneity assumptions, endogenized for the forecasts risking ill-specified relationships, or multi-step estimation has to be adopted. The last of these is the subject of other research and will not be addressed here: see *inter alia* Bhansali (2002) and Chevillon & Hendry (2004).

Moreover, the importance of parameter-estimation uncertainty is heavily dependent on the postulated nature of the DGP and the specific transformations of the variables to be forecast. Specifically, if the DGP is near non-stationary, but treated as stationary, and the horizon is relatively long compared to the estimation sample period available, then parameter-estimation uncertainty plays a fundamental role: see e.g., Stock (1996). The opposite extreme is when the DGP is stationary and a long estimation sample is available, in which case estimation becomes relatively irrelevant as the horizon increases: see the results in Chong & Hendry (1986), noting the potential non-monotonicity of interval forecasts after the first few steps ahead. In between, there is often a surprisingly small component of forecast uncertainty deriving from parameter-estimation uncertainty: see Clements & Hendry (1998), although cases where it matters also occur, as in Marquez & Ericsson (1993).

4.1.1 Estimation Uncertainty

To illustrate the changes which arise in the impact of parameter-estimation uncertainty in dynamic forecasts from dynamic models, we use the first-order autoregression discussed in Clements & Hendry (1998):

$$y_t = \rho y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right],$$

when $|\rho| < 1$, $\mathsf{E}[y_t] = 0$, and $\mathsf{E}\left[y_t^2\right] = \sigma_y^2 = \sigma_\epsilon^2 / \left(1-\rho^2\right)$. Projecting h-steps ahead:

$$y_{T+h} = \rho^h y_T + \sum_{j=0}^{h-1} \rho^j \epsilon_{T+h-j}, \tag{53}$$

so the conditional forecast error from $\widehat{y}_{T+h|T} = \widehat{\rho}^h y_T$ is:

$$\widehat{\epsilon}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T} = \left(\rho^h - \widehat{\rho}^h\right) y_T + \sum_{j=0}^{h-1} \rho^j \epsilon_{T+h-j}. \tag{54}$$

On a MSFE measure where $y_T^\dagger = y_T/\sigma_y$:

$$\mathsf{M}_{AR(1)}\left[\widehat{\epsilon}_{T+h|T} \mid y_T\right] \simeq \sigma_y^2 \left[\left(1-\rho^{2h}\right) + T^{-1} h^2 \rho^{2(h-1)} \left(1-\rho^2\right) y_T^{\dagger 2}\right]. \tag{55}$$

The first term is due to the error variance accumulation as the horizon grows, and the second to parameter-estimation uncertainty. As is well known, overall (55) is not monotonic in the horizon, as the second term tends to increase first before converging to zero. Unconditionally, averaging across $y_T^{\dagger 2}$ yields:

$$\mathsf{M}_{AR(1)}\left[\widehat{\epsilon}_{T+h|T}\right] \simeq \frac{\sigma_\epsilon^2\left(1-\rho^{2h}\right)}{\left(1-\rho^2\right)} + \frac{\sigma_\epsilon^2}{T} h^2 \rho^{2(h-1)}. \tag{56}$$

4.1.2 Simplification

One practical alternative to estimating an unknown parameter is to restrict it to zero, such that the corresponding regressor is omitted. Consider the competing forecast based on omitting the lagged dependent variable, so the forecast becomes the unconditional mean of zero, $\widetilde{y}_{T+h} = 0 \ \forall h$ with:

$$\mathsf{M}_{AR(0)}\left[\widetilde{\epsilon}_{T+h|T} \mid y_T\right] = \sigma_y^2 \left[\left(1-\rho^{2h}\right) + \rho^{2h} y_T^{\dagger 2}\right]. \tag{57}$$

The first terms in (55) and (57) are the same, so the relative MSFE difference, denoted $\mathsf{R}\left(\widetilde{\epsilon}, \widehat{\epsilon}, h\right)$, is:

$$\begin{aligned} \mathsf{R}\left(\widetilde{\epsilon}, \widehat{\epsilon}, h\right) &= \frac{\mathsf{M}_{AR(0)}\left[\widetilde{\epsilon}_{T+h|T} | y_T\right] - \mathsf{M}_{AR(1)}\left[\widehat{\epsilon}_{T+h|T} | y_T\right]}{\sigma_y^2} \\ &= T^{-1} \rho^{2(h-1)} \left(1-\rho^2\right) \left[\tau_{\rho=0}^2 - h^2\right] y_T^{\dagger 2}, \end{aligned} \tag{58}$$

where:

$$\tau_{\rho=0}^2 = \frac{T\rho^2}{1-\rho^2}$$

is the non-centrality of the t-test of H_0: $\rho = 0$ in the DGP equation. Thus, $\tau^2_{\rho=0} > h^2$ is necessary in (58) for an improvement over simply using the unconditional mean. For a 1-step forecast, the criterion is simply $\tau^2_{\rho=0} > 1$. Notice that $100\mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, h)$ is the loss/gain as a percentage of σ^2_y.

As an illustration, if $T = 20$ and $\rho = 0.4$, then $\tau^2_{\rho=0} \simeq 3.8$ so for $y^{\dagger 2}_T = 1$:

$$\mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 1) = 0.12; \ \ \mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 2) = -0.001; \ \ \mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 3) = -0.006;$$

whereas when $\rho = 0.8$, $\tau^2_{\rho=0} \simeq 35.6$ with:

$$\mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 1) = 0.62; \ \ \mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 2) = 0.36; \ \ \mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 3) = 0.20; \ \ \mathsf{R}(\widetilde{\epsilon}, \widehat{\epsilon}, 12) = -0.01,$$

and the sign reverses at $h = 6$. For larger ρ, the sign reverses even later. Once the sign changes, however, the percentage loss stays small, so it is irrelevant if either the conditional or unconditional forecast is chosen for longer horizons.

We conclude that for the simple AR(1) model:

1. there is a clear and measurable trade-off between the costs of estimation and those of omission;

2. the trade-off relates directly to the significance of the variable in the DGP equation via the non-centrality of the t-test;

3. the costs are $\mathsf{O}(T^{-1})$, but could nevertheless be large;

4. the trade-off criterion becomes more stringent as the forecast horizon increases; but

5. once the costs are balanced at some horizon, they stay small for longer horizons.

Points 1.–5. seem to suggest selecting different models at different horizons. However, the last point is crucial for model selection: even relatively insignificant estimates should contribute to forecasting (i.e., variables with $\tau^2_{\rho=0} > 1$). Replicating this finding in more general settings suggests that one need not worry greatly about point 4. In other words, provided $\tau^2_{\rho=0} > 1$, then even if $\tau^2_{\rho=0} < 4$ say, there will be a gain at 1-step, and little additional loss for horizons beyond 4, so the advantages of switching specification after $h = 1$ are unlikely to be large. Nevertheless, checking on the properties of multi-step estimation in this context for more general models would be worthwhile. However, the VAR setting considered by Clements & Hendry (1998) did not deliver any clear recommendations.

5 Empirical results: Euro area inflation

In this section, we analyze empirically the following questions: First, does including the disaggregate variables in the aggregate model improve the direct forecast of the aggregate? Second, is including disaggregate information in the aggregate model better in terms of forecast accuracy than forecasting disaggregate variables and aggregating those forecasts? Third, does it improve the indirect forecast of the aggregate to include aggregate information in the component models? Fourth, does including additional macro-economic predictors improve the aggregate forecast? We relate the findings to the predictability results from section 2 and the analytical results regarding the effects of model selection and estimation from section 3. Whether we find that the result from a general theory of prediction that including disaggregate variables does improve predictability of the aggregate will hold empirically, will depend on the effect of model selection and estimation, i.e., on the trade-off between improving the forecast accuracy of the mean by retaining a variable in the model on the one hand and adding to the forecast error variance on the other hand. In the context of forecasting an aggregate by disaggregates this will depend on how collinearity and component weights change over the forecast period. However, all the theoretical implications above assume the absence of other complicating factors, such as location shifts and measurement errors, which may play an important role in practice.

5.1 Data

The data employed in this study include aggregated overall HICP for the Euro area as well as its breakdown into five subcomponents: unprocessed food, processed food, industrial goods, energy and services prices.

This particular breakdown into subcomponents has been chosen in accordance with the data published and analyzed in the ECB Monthly Bulletin. A range of explanatory variables for inflation is also considered: Industrial production, nominal money M3, producer prices, import prices (extra euro area), unemployment, unit labour costs, commodity prices (excluding energy) in euro, oil prices in euro, the nominal effective exchange rate of the euro[8], as well as a short-term and a long-term nominal interest rate.

The data employed are of monthly frequency,[9] starting in 1992(1) until 2001(12). This relatively short sample is determined by the availability of data for the Euro area and has to be split for the out-of-sample forecast experiment. Seasonally ad-

[8]ECB effective exchange rate core group of currencies against euro.

[9]Except for unit labour costs which are of quarterly frequency and have been interpolated.

justed data have been chosen[10] because of the changing seasonal pattern in some of the HICP subcomponents for some countries due to a measurement change.[11]

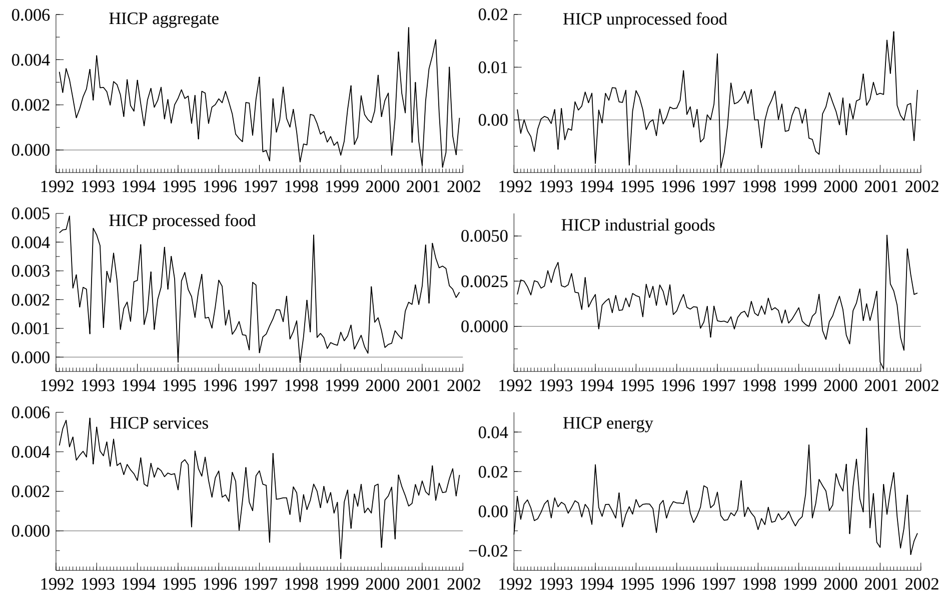

Figure 1: First differences of HICP (sub-)indices (in logarithm)

The month-on-month inflation rates (in decimals) and the year-on-year inflation rates (in %) of the indices are displayed in Figures 1 and 2, respectively.

We have carried out Augmented Dickey Fuller (ADF) tests for all HICP (sub-) indices (in logarithms), since Diebold & Kilian (2000) show for univariate models that testing for a unit root can be useful for selecting forecasting models. The tests are based on the sample from 1992(1) to 2000(12). This is the longest of the recursively estimated samples in the simulated out-of-sample forecast experiment (see Section 5.3). The tests do not reject non-stationarity for the levels of all (sub-) indices over the whole period.[12] Non-stationarity is rejected for the first differences of all series except the aggregate HICP and HICP services. For the first differences of the latter two series, however, non-stationarity is rejected for all shorter recursive

[10]Except for interest rates, producer prices and HICP energy that do not exhibit a seasonal pattern.

[11]The data used in this study are taken from the ECB and Eurostat.

[12]The ADF test specification includes a constant and a linear trend for the levels and first differences. The number of lags included is chosen according to the largest significant lag on a 5% significance level.

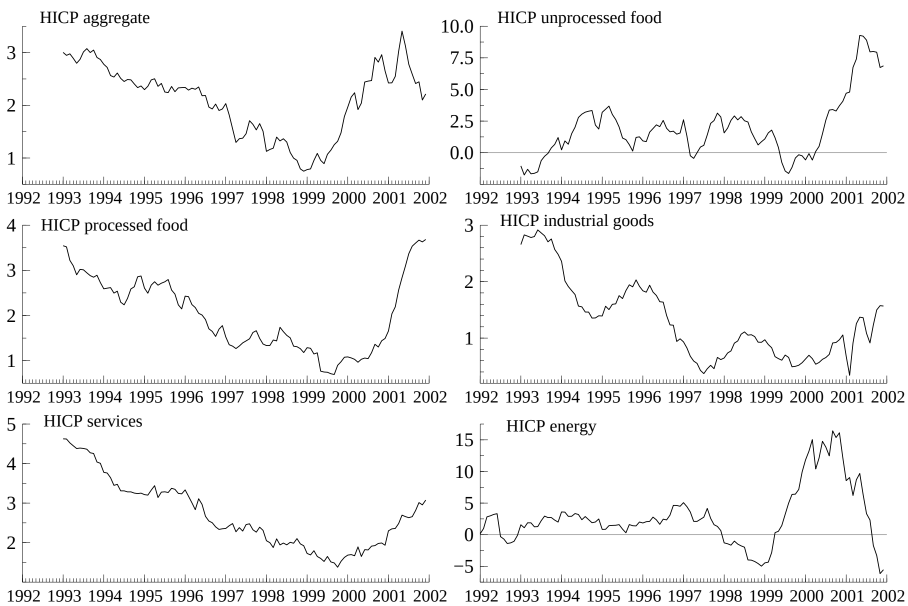

Figure 2: Year-on-year HICP inflation (in %), aggregate and subindices

estimation samples up to 2000(8) and 2000(7), respectively. Therefore and because of the low power of the ADF test HICP (sub-)indices are assumed to be integrated of order one in the analysis and modeled accordingly.

5.2 Forecast methods and model selection

Different forecasting methods using different model selection procedures are employed for both direct and indirect forecast methods, i.e., forecasting HICP inflation directly versus aggregating subcomponent forecasts. We employ simple autoregressive (AR) models where the lag length is selected by the Schwarz (SIC) and the Akaike (AIC) criterion respectively (see e.g. Inoue & Kilian, 2005). We include a subcomponent vector autoregressive model (VAR^{subc}) to indirectly forecast the aggregate by aggregating subcomponent forecasts. We use a VAR including the aggregate and the components, VAR^{agg,sub}, to investigate the hypothesis from section 2 that including component information in the aggregate forecast model improves the forecast of the aggregate. We include a VAR where the lags of the aggregate and the components are automatically chosen using *PcGets*, VAR$^{agg,sub}_{Gets}$ (see Hendry & Krolzig, 2003). In a second group of methods, we include additional macroeconomic predictors in the VARs of the aggregate and the components, respectively. This group includes two VARs with a set of domestic and international variables, the VAR^{int}, where the specification is the same across the aggregate and the components, and the VAR$^{int}_{Gets}$, where the specification is allowed to vary across components. Finally, a VAR including potentially all variables, i.e., aggregate, components and other macroeconomic predictors, VAR$^{IntAggSub}_{Gets}$, is considered.[13] The lag length of the VAR is selected on the basis of the SIC, the AIC and an F-test.[14]

5.3 Simulated out-of-sample forecast comparison

5.3.1 The experiment

A simulated out-of-sample forecast experiment is carried out to evaluate the relative forecast accuracy of alternative methods to forecast aggregate HICP using information on its disaggregate components as opposed to aggregating the forecasts

[13] For the forecast accuracy results presented in the tables of this section model selection procedures are carried out on the basis of the first recursive estimation sample until 1998(1). However, recursive model selection was carried out for the most relevant models. The results did not suggest a change to our conclusions.

[14] It should be noted that due to the large number of parameters in the high-dimensional VARs the maximum lag order was chosen on the basis of a rough rule such that the total number of parameters in the system would not exceed half the sample size.

of HICP subcomponent models or forecasting the aggregate only using aggregate information. One to twelve step ahead forecasts are performed based on different linear time series models estimated on recursive samples. The main criterion for the comparison of the forecasts employed in this study, as in a large part of the literature on forecasting, is the root mean square forecast error (RMSFE).

Table 2 and 5 present the comparison of the relative forecast accuracy measured in terms of RMSFE of year-on-year (headline) inflation of the direct forecast of aggregate inflation ($\Delta_{12}\widehat{p}^{agg}$) and the indirect forecast of aggregate inflation, i.e. the aggregated forecasts of the sub-indices ($\Delta_{12}\widehat{p}^{agg}_{sub}$). The results for 1-,6- and 12-months ahead forecasts are presented.

5.3.2 Aggregate and disaggregate information

First we compare methods only based on aggregate information as opposed to forecast methods for the aggregate including disaggregate variables in addition (see Table 2, column for direct forecast for each forecast horizon). Within the framework of the general theory of prediction we have shown that including disaggregate variables in the aggregate model does improve predictability of a variable (see section 2). We find that the direct forecast using a VAR including the aggregate and subcomponents where the variables are selected by PcGets, VAR$^{agg,sub}_{Gets}$, performs slightly better in RMSFE terms 1 month ahead than directly forecasting the aggregate with an AR model only including lagged aggregate information with the lag length determined by the SIC criterion. Thus, our RMSFE results for the VAR$^{agg,sub}_{Gets}$ for $h = 1$ confirm this predictability result in a forecast experiment. However, the model including the aggregate and all subcomponents, VAR$^{agg,sub}_{(1)}$ does not provide a more accurate forecast of the aggregate than the autoregressive models AR^{SIC} and AR^{AIC}.

Furthermore, we investigate the accuracy of forecasting the aggregate directly including disaggregate variables relative to the forecast accuracy of indirectly forecasting the aggregate by aggregating component forecasts based on an AR model or a subcomponent VAR, VAR^{sub}, (Table 2), i.e., the way previous literature has been taken disaggregate variables into account (see e.g. Hubrich (2004)). The VAR model that outperforms the other direct forecast methods of the aggregate, VAR$^{agg,sub}_{Gets}$, also exhibits higher forecast accuracy for the indirect forecast than all other methods for $h = 1$. Thus, including aggregate variables in the disaggregate model improves forecast performance for short horizons. The VAR$^{agg,sub}_{Gets}$ does also outperform the VAR$^{agg,sub}_{(1)}$, where the variables and lag length are the same across the aggregate and components, for $h = 1$.

Overall, the direct forecast including the aggregate and subcomponents is best for 1 month ahead forecasts if no additional macroeconomic indicators are con-

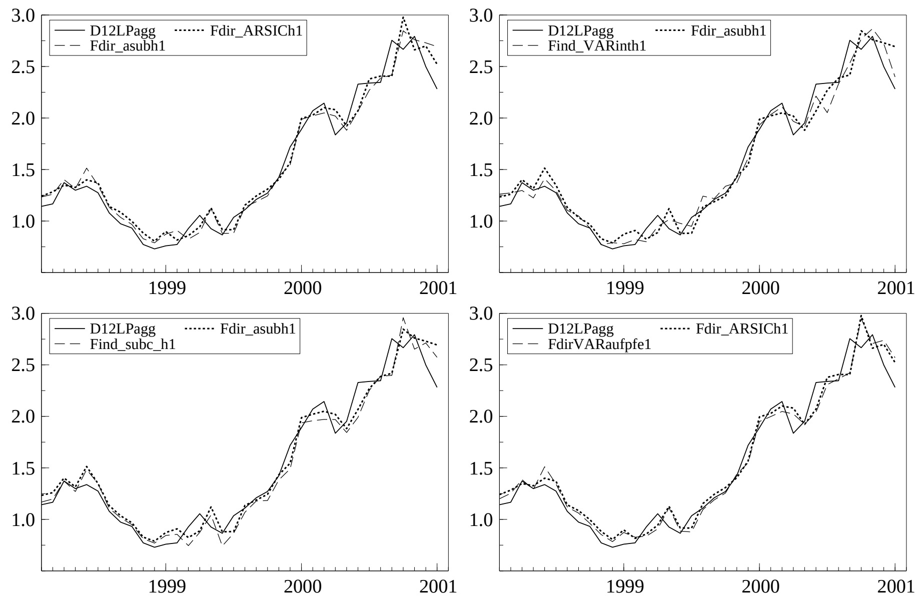

Figure 3: Year-on-year inflation rate and forecasts in %, 1 months ahead, solid line: actual, Fdir: direct forecast of aggregate, Find: indirect forecast of aggregate

sidered. That confirms within a forecasting set-up the results derived with respect to predictability in section 2, i.e. that forecasting the aggregate directly including disaggregate information in the aggregate model might perform better than aggregating component forecasts. Figure 3 shows that the one months ahead forecasts from the different methods are very close to actual year-on-year inflation. The differences between the different methods for one month ahead forecasts appear to be quite small. Figure 4 presents the forecast 6 months ahead. Six months ahead forecasts of year-on-year inflation do generally relatively well. The graphs show that the differences in RMSFE terms between some of the forecasts are relevant to be considered when choosing the forecasting model.

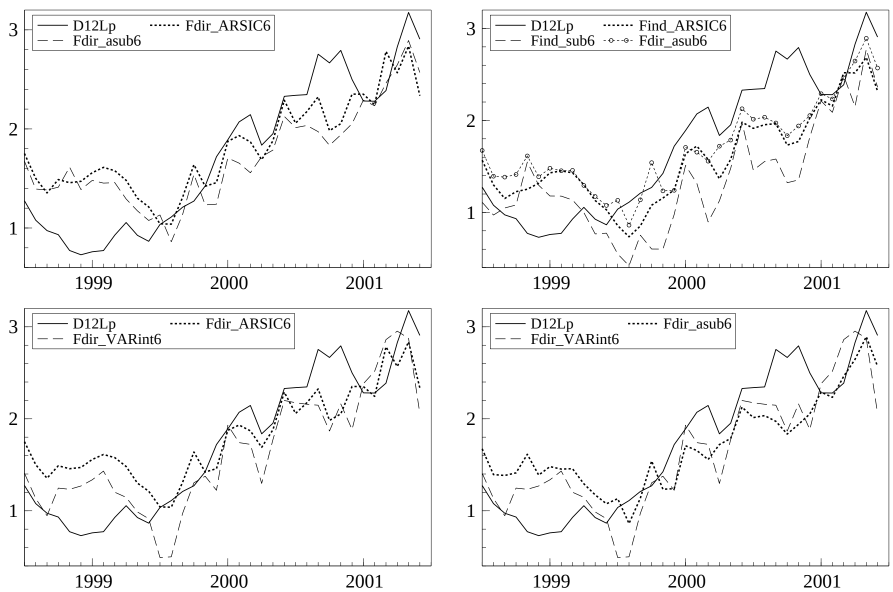

Figure 4: Year-on-year inflation rate and forecasts in %, 6 months ahead, solid line: actual, Fdir: direct forecast of aggregate, Find: indirect forecast of aggregate

More important from a monetary policy point of view is the 12 months ahead forecast. Here we find that the direct forecast including disaggregate information ($\text{VAR}_{(1)}^{agg,sub}$) is clearly better than the indirect forecast based on AR or VAR models of the components. The low forecast accuracy of aggregating subcomponent models is analyzed in Hubrich (2004), and it is found that this is due to unexpected shocks that occur in the forecast period and affect some or all components in the same direction so that forecast errors do not cancel. Furthermore, predictability in the sense we have defined in section 2.1 is low for some component series and their unconditional variance is large. Consequently they are very difficult to forecast. This leads to low forecast accuracy of the indirect forecast of the aggregate. Hubrich (2004) investigates whether forecast combination of different methods improves forecast accuracy of the components and hence the indirect forecast of the aggregate and finds that this is not the case. However, directly forecasting the aggregate using VAR$^{agg,sub}_{(1)}$ is very similar in terms of forecast accuracy to using the same model, i.e. including all disaggregate variables and the aggregate, for the indirect forecast. Including the aggregate in the component models seems to improve forecast accuracy of the aggregate.[15] We find that the indirect forecast based on a VAR including subcomponents with no lags as chosen by the SIC exhibits higher forecast accuracy than all other indirect forecast methods. This model represents a random walk with drift for prices for each of the components and for the aggregate and is selected by the SIC for the VAR^{sub}, the $VAR^{agg,sub}$ and the VAR^{int}. However, the direct forecast using a simple AR model does lead to the highest forecast accuracy 12 months ahead overall.

For a 12 months ahead horizon the VAR$^{agg,sub}_{(1)}$ outperforms the VAR$^{agg,sub}_{Gets}$ in contrast to the one-month ahead result. Furthermore, note that the AR^{SIC} performs better than the AR^{AIC} for one and twelve months ahead forecasts in line with the results in section 4.1 that the trade-off between costs of estimation and omission becomes more stringent as the forecast horizon increases (see also Inoue & Kilian (2005) for a comparison of the SIC and the AIC).

Although perfect collinearity between aggregate and components does not pose a problem due to annually changing weights in price indices, we present additional results where different subsets of price components are selected. Comparing forecast accuracy of the VAR^{agg,i,s,pf}, the VAR^{agg,e,uf,pf} and the VAR^{agg,e,uf} we find that selection from disaggregate variables seems to improve the forecast accuracy. When we exclude processed food inflation from the VAR^{agg,e,uf,pf}, the lower dimensional VAR^{agg,e,uf} does perform worse than the VAR^{agg,e,uf,pf}, in particular for $h = 12$. Therefore, the decline in collinearity of the excluded variable and the variables in the system does matter for the forecast accuracy of the method (see also section 3).

Figure 5, in the upper two panels, shows the relevance of the differences in the forecasts for h=12 discussed above from a monetary policy viewpoint, i.e., a difference between 0.2 up to almost 2 percentage points for some methods in some periods of the forecast period is clearly relevant in this context. Tests to compare

[15] It should be noted, however, that this result depends on the lag length of 2 suggested by the F-test. A lag order of zero as chosen by the SIC, representing a random walk with drift for prices, provides a more favourable result for the indirect forecast with no aggregate included in the component models.

Table 2: **Aggregate and disaggregate information, Relative forecast accuracy: Average RMSFE ratios over AR(p) of year-on-year inflation in percentage points**

horizon	1		6		12	
method	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$
AR^{SIC}	0.137	0.139	0.431	0.478	0.740	0.880
AR^{AIC}	1.015	1.014	0.937	1.000	1.245	1.027
VAR$^{sub}_{(2)}$		1.086		1.402		1.593
VAR$^{agg,sub}_{(0)}$	1.007	0.993	1.039	0.937	1.035	0.876
VAR$^{agg,sub}_{(1)}$	1.044	1.036	1.046	0.939	1.081	0.9l3
VAR$^{agg,sub}_{Gets}$	0.978	0.964	1.079	0.960	1.124	0.933
VAR$^{agg,i,s,pf}_{(1)}$	1.029		1.030		1.042	
VAR$^{agg,e,uf,pf}_{(1)}$	1.007		1.009		1.022	
VAR$^{agg,e,uf}_{(1)}$	1.007		1.019		1.030	

Note: RMSFE displayed for $AR^{(}SIC)$ model, Recursive estimation samples 1992(1) to 1998(1),...,2000(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, AIC: Akaike criterion, VAR^{sub}: VAR only including subcomponents, lag order, $p = 2$, VAR^{agg,sub}: VAR with aggregate and subcomponents $p = 0$ (SIC), $p = 1$ (AIC), VAR$^{agg,sub}_{Gets}$: VAR with aggregate and subcomponents selected by PcGets, liberal strategy Hendry & Krolzig (2001)

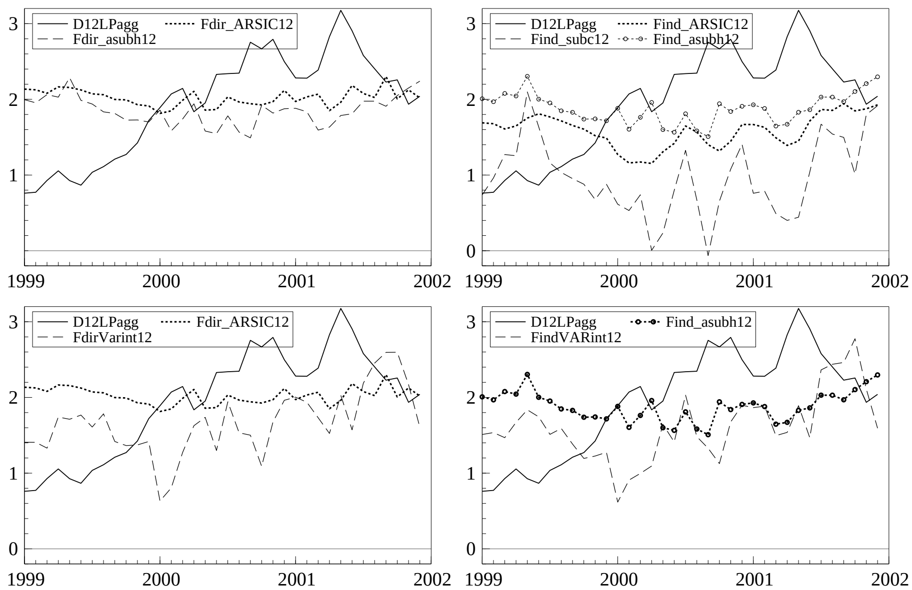

Figure 5: Year-on-year inflation rate and forecasts in %, 12 months ahead, solid line: actual, Fdir: direct forecast of aggregate, Find: indirect forecast of aggregate

Table 3: **Factor models based on disaggregate prices only, RMSFE and average RMSFE ratios of direct forecast of annualised inflation in percentage points**

horizon	1	6	12
RMSFE AR^{SIC}	0.180	0.396	0.780
RMSFE ratios over AR^{SIC}			
$FM(f1)$	1.036	1.101	1.084
$FM(f2)$	1.016	1.078	1.046
$FM(f3)$	1.024	1.097	1.067
$FM(f4)$	1.044	1.086	1.105
$DFM(f1)^{SIC}$	1.117	1.157	1.390
$DFM(f2)^{SIC}$	1.013	1.110	1.007
$DFM(f3)^{SIC}$	1.029	1.225	0.955
$DFM(f4)^{SIC}$	1.062	1.333	1.116

Note: RMSFE for AR$^{(SIC)}$ model, Recursive estimation samples 1992(1) to 1998(1),...,2000(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, FM(f): factor models with 1,2,3,4 factors, DFM(f)^{SIC}: dynamic factor models with 1,2,3,4 factors with factor lag lengths chosen by SIC

the significance of the difference in forecast accuracy are not carried out due to their poor size (and power) properties in small forecast samples as considered here (for simulation evidence see e.g. Harvey, Leybourne & Newbold (1997)).

In section 3.4 we have analysed theoretically the effects of different types of changes influencing forecast accuracy of the aggregate model including disaggregate components. We now analyse two of those changes in the context of forecasting euro area inflation: a change in component weights and a change in collinearity of disaggregate regressors.

There is some change in (consumer spending) weights of euro area price components: Weights decline between -3.9 % and -1.3 % annually on average over the previous year over the forecast evaluation period for unprocessed food, processed food and industrial goods prices, where in one year for example the decline is almost -9% for unprocessed food. For energy prices weights decline by -6.5% in 1999 and then increase by 3.4% and 5.6 %, respectively. Service price weights increase by 3% on average per year over the forecast evaluation period. These changes in weights mean that the relevance of the changes of, say, unprocessed food prices for the aggregate declines over the forecast evaluation period so that positive shocks to unprocessed food prices does affect the aggregate less, whereas the positive shocks to energy prices will affect the aggregate more in the future. [16]

Second, we analyse the change in the correlation structure between the aggregate and the components over the forecast evaluation period. Table 4 presents the correlation matrix, where the upper triangle represents the correlation for the first estimation sample until 1998(1) and the lower triangle represents the correlation for the last forecast sample up to 2001(12). Most of the time correlations between aggregate and components, particularly large declines are observed between Δp^{agg} and Δp^{s}. Overall, correlations between the aggregate and the components decline. Including the respective component(s) in the forecast model might then lower forecast accuracy by increasing estimation uncertainty. This might help explaining that selection pays according to the results in Table 2 where the $\text{VAR}^{agg,sub}_{Gets}$ outperforms all other models one month ahead. Furthermore, correlation among disaggregate components included in the models decline, i.e. collinearity is lower between the regressors. This will affect forecast accuracy as discussed in section 3. A particularly large decline in correlation can be found in Δp^{uf} and Δp^{pf} as well as Δp^{i} and Δp^{s}. In some cases even the sign switches: Δp^{uf} and Δp^{pf}, Δp^{uf} and Δp^{i} as well as Δp^{e} and Δp^{s}, with low negative correlation between those components for the longest sample. This increases the costs of omission of the respective components, as is apparent in the lower forecast accuracy of the more parsimonious model $\text{VAR}^{agg,e,uf}_{(1)}$ in comparison with $\text{VAR}^{agg,e,uf,pf}_{(1)}$.

The above effects favour an aggregate model, in particular for longer forecast horizons as a year, in the sense that an aggregate only including lags of the aggregate might be a more robust forecasting device when the effect of changing weights and collinearity on the trade-off between the costs of estimation and those of omission is unknown a priori.

5.3.3 Extending the information set further: Macroeconomic Predictors

In Table 5 the RMSFE results for models including additional macroeconomic predictors are presented alongside the most relevant methods only including aggregate and disaggregate information. Including relevant macroeconomic predictors changes the findings. In that case, the indirect methods tend to perform better one month ahead and the international VAR (VAR^{int}) does outperform all other indirect forecasts and also the direct forecasts for $h = 1$ and $h = 6$. Even a VAR method that initially allows for all variables to enter the forecast model where the relevant variables are selected (VAR^{Int}_{Gets}) is outperformed by the VAR^{int}. This can be explained by our analytical results in section 3 that imply choosing a loose significance level in forecast model selection.

[16] The indirect forecast of the aggregate by aggregating the component forecasts is also affected

Table 4: **Correlation matrix of first differences of log prices: upper triangle sample until 1998(1), lower triangle sample until 2001(12)**

	Δp^{agg}	Δp^{uf}	Δp^{pf}	Δp^{i}	Δp^{s}	Δp^{e}
Δp^{agg}	1	0.27931	0.43579	0.63135	0.70929	0.60797
Δp^{uf}	0.3229	1	-0.22429	-0.1362	-0.13113	0.036102
Δp^{pf}	0.34893	0.028023	1	0.41859	0.52704	0.0048911
Δp^{i}	0.53733	0.044988	0.28777	1	0.61273	0.085604
Δp^{s}	0.49376	-0.06515	0.51041	0.45057	1	0.10458
Δp^{e}	0.7071	0.0076403	-0.078966	0.039373	-0.06116	1

Table 5: **Including macroeconomic predictors, Relative forecast accuracy: Average RMSFE ratios of year-on-year inflation in percentage points**

horizon	1		6		12	
method	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$	direct $\Delta_{12}\widehat{p}^{agg}$	indirect $\Delta_{12}\widehat{p}^{agg}_{sub}$
AR^{SIC}	0.137	0.139	0.431	0.478	0.740	0.880
VAR$^{sub}_{(2)}$		1.086		1.402		1.593
VAR$^{agg,sub}_{(1)}$	1.044	1.036	1.046	0.939	1.081	0.913
VAR$^{agg,sub}_{Gets}$	0.978	0.964	1.079	0.960	1.124	0.933
VAR$^{Int(2)}$	0.839	0.748	0.979	0.937	1.065	0.964
VAR$^{Int}_{Gets}$	1.022	0.842	1.186	1.000	1.272	0.966
VAR$^{IntSub}_{Gets}$	1.015	1.007	1.188	1.073	1.219	1.031

Note: RMSFE for $AR^(SIC)$ model, Recursive estimation samples 1992(1) to 1998(1),...,2000(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, VAR^{sub}: VAR only including subcomponents, lag order $p = 2$, VAR^{agg,sub}: VAR with aggregate and subcomponents $p = 0$ (SIC), $p = 1$ (AIC), VAR$^{agg,sub}_{Gets}$: VAR with aggregate and subcomponents selected by PcGets, liberal strategy Hendry & Krolzig (2001) VAR$^{Int(p)}$: model including international and domestic variables, lag length $p = 2$, VAR$^{Int}_{Gets}$: variables as VAR^{Int}, model selection with PcGets, VAR$^{IntSub}_{Gets}$: as VAR^{Int}, additionally including subcomponents, model selection with PcGets,

Table 6: **Including macroeconomic predictors, RMSFE and average RMSFE ratios of direct multi-step forecast of annualised inflation in percentage points**

horizon	1	6	12
RMSFE AR^{SIC}	0.180	0.396	0.780
RMSFE ratios over AR^{SIC}			
$FM(f1)$	1.016	0.970	1.005
$FM(f2)$	0.966	0.943	0.973
$FM(f3)$	0.943	0.968	0.961
$FM(f4)$	0.877	0.981	0.948
$DFM(f1)^{SIC}$	0.986	0.986	1.015
$DFM(f2)^{SIC}$	0.980	0.881	0.972
$DFM(f3)^{SIC}$	0.997	1.172	0.890
$DFM(f4)^{SIC}$	0.917	1.313	1.065

Note: Recursive estimation samples 1992(1) to 1998(1),...,2000(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, FM(f): factor models with 1,2,3,4 factors, DFM(f)^{SIC}: dynamic factor models with 1,2,3,4 factors with factor lag lengths chosen by SIC

The 12 months ahead forecast of the VAR^{int} is more accurate than the VAR$^{agg,sub}_{(1)}$ model including the aggregate and its subcomponents. For this horizon the VAR^{agg,sub} outperforms VAR$^{agg,sub}_{Gets}$. However, the AR model is now performing best overall, even better than the VAR^{int}. Regarding the indirect methods the VAR including aggregate information instead of macroeconomic predictors performs best. The lower two panels of figure 5 show the relevance of the differences in the forecasts for h=12 when additional macroeconomic predictors are included in the models from a monetary policy viewpoint.

Overall, including additional macroeconomic predictors improves the forecast accuracy of directly forecasting aggregate inflation over all other linear models considered except for a forecast horizon of 12 where the AR model outperforms all the others. This might be attributed to the change in collinearity between the explanatory variables which will influence the trade-off in forecast model selection between retaining and omitting variables (see section 3). Examining the change in correlations between the variables included in the VAR^{int} reveals that the collinearity is changing over time, with weaker (positive or negative) correlation between many variable for the last recursive estimation sample up to 2000(12) than in the first sample up to 1998(1). This provides a rationale why the AR^{SIC} might be bet-

since the weights are used for aggregation.

ter for longer forecast horizons than the VAR^{int} whereas the opposite is found for short forecast horizons. For long horizons, the change in collinearity is dominating the result from section 4.1 for the dynamic forecast that for large h model selection does not necessarily have to be parsimonious and the specification that is best for short horizons might be retained. A parsimonious model is to be preferred due to the change in collinearity.[17]

For the indirect forecast macroeconomic predictors improve forecast accuracy for a 1 month horizon whereas for a 12 months horizon including aggregate information in the disaggregate model is better. Figure 5, the lower two panels, shows that differences in RMSFE terms between the direct forecasts of the AR and the international VAR model matter for 12 step ahead forecasts.

The out-of-sample forecasting experiment suggests that for 1 month ahead forecasts the information set selected is an important determinant of the forecast accuracy. In an iterative multi-step forecasting set-up the parsimony of the forecast model specification appears to be a very important factor for longer forecast horizons, in particular in the relatively small sample available and under changing collinearity and structural breaks. In this case estimation uncertainty of a high number of parameters increases the MSFE relatively more.

6 Conclusions

In this paper, we show that a theory of prediction suggests that including disaggregate variables in an aggregate model should outperform in terms of predictability an aggregate model which only includes lags of the aggregate. However, we find that it does not always do so when forecasting euro area inflation.

There are many steps between predictability in population and 'forecastability' where the forecast model might differ from the data generation process. Recall that the predictability concept that we consider in this paper refers to a property of the variable of interest in relation to the information set considered. In contrast, forecastability refers to the improvement in forecast accuracy given the unconditional moments of a variable based on the information set available. The predictive value of disaggregate information can be off-set by estimation uncertainty; model selection; changing collinearity, as measured by the ratio of eigenvalues λ^*/λ; and unmodeled breaks. The effect of estimation uncertainty is O(T^{-1}) or O$(T^{-1}\lambda^*/\lambda)$ depending on whether collinearity is unchanged or changing in the forecast period.

[17] Recursive model selection has been carried out for the AR^{SIC} and the VAR^{Int}_{Gets} and the forecast accuracy has been compared with the methods considered so far. The forecast accuracy did not improve over the AR^{SIC} and the VAR^{Int} presented in the tables and hence did not change our conclusions.

For model selection, a loose significance level is suggested based on the analytical results for a forecast horizon of one period in section 3. Collinearity is shown to be irrelevant if it is unchanged, but changes in collinearity of both included and incorrectly-excluded variables affect forecasts. Section 4.1 suggests that for longer horizons, selection of an even more parsimonious model may not improve forecast accuracy greatly. However, which method performs best in terms of forecast accuracy would then depend on other factors, such as changes in collinearity. Furthermore, the theoretical implications of predictability assume the absence of other complicating factors, such as location shifts and measurement errors, which may play an important role in practice.

In the context of forecasting euro area inflation, changing weights in the price index and changing collinearity between disaggregate prices both act against disaggregate-based models. When only considering aggregate and disaggregate variables as the initial information set included in the forecast model, the following conclusions can be drawn from our empirical analysis: Overall, we find that there is little cost or benefit from selection in multivariate models for short horizons, although the model chosen by *PcGets* is best at a forecast horizon of $h = 1$. However, more stringent selection pays as h grows when comparing AR models based on the SIC versus AIC. Indirect forecasts, i.e., forecasting the disaggregates and aggregating those forecasts, usually perform worst, although the selection procedure does play a role. Furthermore, including aggregates as regressors in the VAR of disaggregate variables might pay, again depending on the lag order selection procedure applied. All 1-month ahead forecasts are quite close, whereas the differences between the different methods increase as h grows. Overall, the theoretical result on predictability that more disaggregate information does help does not find strong support in this forecasting context. Dynamic factor forecasts, where the factors are derived based on disaggregate price variables only, improve over the AR model only if 3 factors are included.

If the information set is extended further, including macroeconomic predictors, some of the previous results change. We still find that more stringent selection is worthwhile as h grows since the AR model selected by *SIC* clearly outperforms the other methods in terms of forecast accuracy. Similar to including disaggregate information, additional information in terms of macroeconomic predictors helps for shorter forecast horizons. The large 'international' VAR exhibits higher forecast accuracy than the AR model based on the *SIC* for short forecast horizons. Furthermore, there is little benefit from selection when comparing the large 'international' VAR with the VAR selected by *PcGets*. There appears to be no benefit from additionally including disaggregates in the international VAR. Dynamic factor models do also improve over the AR model in terms of forecast accuracy.

All methods perform quite similar in terms of forecast horizon of one-step ahead. For a forecast horizon of one year, differences between the forecasts from the different methods are larger and are relevant from a monetary policy perspective.

We can now answer the four questions posed above for forecasting euro area inflation and our analytical results help explain those empirical results. First, including disaggregate variables in the aggregate model does not really improve forecasts of the aggregate, in particular for longer forecast horizons. Second, including disaggregate information in the aggregate model might be better than forecasting disaggregates and aggregating those forecasts. Third, including aggregate information in component models might improve those forecasts of the aggregate that are derived by aggregating component forecasts. Fourth, including additional macro-economic predictors does improve the aggregate forecast for shorter forecast horizons, but not for longer.

That disaggregate variables and macro-economic predictors do not necessarily improve forecast accuracy in our empirical application seems to be in contrast to our theoretical results in the context of predictability in population. However, it might be attributed to a large extent to changing collinearity between components and macro-economic predictors as regressors in the aggregate model. Since the effect of changing collinearity on the model selection trade-off between the costs of estimation and those of omission is unknown in practice, an aggregate model only including lags of the aggregate might be considered a more robust forecasting device for longer forecast horizons.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov & F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Allen, P. G. & Fildes, R. A. (2001). Econometric forecasting strategies and techniques, *in* J. S. Armstrong (ed.), *Principles of Forecasting*, Kluwer Academic Publishers, Boston, pp. 303–362.

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

Bhansali, R. J. (2002). Multi-step forecasting, *in* M. Clements & D. F. Hendry (eds), *A companion to economic forecasting*, Blackwells, Oxford, pp. 206–221.

Campos, J., Hendry, D. F. & Krolzig, H.-M. (2003). Consistent model selection by an automatic gets approach, *Oxford Bulletin of Economics and Statistics* **65**: 803–819.

Chevillon, G. & Hendry, D. F. (2004). Direct multi-step estimation for forecasting economic processes, *International Journal of Forecasting* pp. –. Forthcoming.

Chong, Y. Y. & Hendry, D. F. (1986). Econometric evaluation of linear macroeconomic models, *RESTUD* **53**: 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

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

Clements, M. P. & Hendry, D. F. (1998). *Forecasting Economic Time Series*, Cambridge University Press, Cambridge, UK.

Clements, M. P. & Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge,Massechusetts.

Clements, M. P. & Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition, *International Journal of Forecasting* **17**: 550–554.

Clements, M. P. & Hendry, D. F. (2004a). Forecasting with breaks, *in* C. W. J. G. G. Elliott & A. Timmermann (eds), *Handbook of Forecasting, forthcoming*, Elsevier.

Clements, M. P. & Hendry, D. F. (2004b). Pooling forecasts, *Econometrics Journal, forthcoming*.

Diebold, F. X. & Kilian, L. (2000). Unit-root tests are useful for selecting forecasting models, *Journal of Business & Economic Statistics* **18**(3): 265–272.

Diebold, F. X. & Kilian, L. (2001). Measuring predictability: Theory and macroeconomic applications, *Journal of Applied Econometrics* **16**: 657–669.

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

Fair, R. C. & Shiller, J. (1990). Comparing information in forecasts from econometric models, *The American Economic Review* **80**(3): 375–389.

Fildes, R. & Ord, K. (2002). Forecasting competitions - their role in improving forecasting practice and research, *in* M. P. Clements & D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 322–253.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000). The generalized facotr model: Identification and estimation, *Review of Economics and Statistics* **82**: 540–554.

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

Harvey, D., Leybourne, S. & Newbold, P. (1997). Testing the equality of prediction mean squared errors, *International Journal of Forecasting* **13**: 281–291.

Hendry, D. F. (1995). *Dynamic Econometrics*, Oxford University Press, Oxford.

Hendry, D. F. (2004). Unpredictability and the foundations of economic forecasting, *Working paper*, Economics Department, Oxford University.

Hendry, D. F. & Krolzig, H.-M. (2001). *Automatic Econometric Model Selection Using PcGets 1.0*, Timberlake Consultants Ltd., London, UK.

Hendry, D. F. & Krolzig, H.-M. (2003). New developments in automatic general-to-specific modelling, *in* B. P. Stigum (ed.), *Econometrics and the Philosophy of Economics*, Princeton University Press, Princeton, pp. 379–419.

Hendry, D. F. & Krolzig, H.-M. (2004). Resolving three 'intractable' problems using a gets approach, *Unpublished paper*, Economics Department, University of Oxford.

Hubrich, K. (2004). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?, *International Journal of Forecasting, in press*.

Inoue, A. & Kilian, L. (2005). On the selection of forecasting models, *Journal of Econometrics, forthcoming*.

Johnson, N. L. & Kotz, S. (1970). *Continuous Univariate Distributions – 2*, John Wiley, New York.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past?, *Journal of Econometrics* (18): 337–349.

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

Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag.

Lütkepohl, H. (2004). Forecasting with VARMA processes, *in* C. W. J. G. G. Elliott & A. Timmermann (eds), *Handbook of Forecasting, forthcoming*, Elsevier.

Makridakis, S. & Hibon, M. (2000). the M3-competition: Results, conclusions and implications, *International Journal of Forecasting* **16**: 451–476.

Marquez, J. R. & Ericsson, N. R. (1993). Evaluating forecasts of the U.S. trade balance, *in* R. Bryant, P. Hooper & C. L. Mann (eds), *Evaluating Policy Regimes: New Research in Empirical Macroeconomics*, Brookings Institution, Washington, DC.

Pesaran, M. H., Pierse, R. G. & Kumar, M. S. (1989). Econometric analysis of aggregation in the context of linear prediction models, *Econometrica* **57**: 861–888.

Stock, J. H. (1996). VAR, error correction and pre-test forecasts at long horizons, *Oxford Bulletin of Economics and Statistics* **58**: 685–701.

Stock, J. H. & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indices, *Journal of Business and Economic Statistics* **20**(2): 147–162.

Van Garderen, K. J., Lee, K. & Pesaran, M. H. (2000). Cross-sectional aggregation of non-linear models, *Journal of Econometrics* **95**: 285–331.