Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate

David F. Hendry and Kirstin Hubrich[1]

Department of Economics, Oxford University
and Research Department, European Central Bank

August 2007

Abstract

The forecast combination literature largely focuses on combining forecasts from different models for the same variable. Instead, our focus is on forecasting an aggregate variable of interest comparing three methods: First, combining forecasts of disaggregate components of the aggregate; second, combining disaggregate information, that is including disaggregate variables in the model of the aggregate; and third, forecasting the aggregate by a univariate model, i.e. only using lagged aggregate information. We present new analytical results on the effects of changing coefficients, mis-specification and estimation uncertainty as well as of changing weights on the relative forecast accuracy of the different methods to forecast an aggregate. Our findings indicate some interesting differences with respect to the usual forecast combination literature. We find that shifts in the mean or slope parameters over the forecast horizon and differences in the variance-covariance structure of the error terms do not affect the relative forecast accuracy of the different approaches to forecast an aggregate. We also find that slope mis-specification and estimation uncertainty are the primary sources of forecast error differences. To reduce estimation uncertainty and thereby the mean square forecast error, variable selection procedures or methods of dimension reduction such as factor models might be employed when including disaggregate information in the aggregate model. We present conditions under which a ranking between the methods to forecast an aggregate is possible. In Monte Carlo simulations we find that different stochastic structures of components and interdependencies between disaggregates also influence the relative forecast accuracy. We find confirmation of our theoretical predictions and simulation results in a pseudo out-of-sample forecast experiment to forecast aggregate pre- and post 1984 US inflation using disaggregate information.

JEL: C51, C53, E31
KEYWORDS: Aggregate forecasts, disaggregate information, forecast combination, inflation

[1]We thank Marcel Bluhm and Eleonora Granziera for excellent research assistance. We are grateful to Domenico Gianonne, Lutz Kilian, Helmut Lütkepohl, Serena Ng, Lucrezia Reichlin, Kenneth West and seminar participants at Bocconi University, Duke University, the Federal Reserve Board, the University of Michigan, the University of Wisconsin and participants of the the EUI Conference on Recent Advances in Econometric Modeling 2006 for useful comments. We thank two anonymous referees and an associate editor for detailed comments and Walter Lane and Patrick Jackman (BLS) for providing the historical weights for the US CPI data. 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 macroeconomic aggregates are employed 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, over simply forecasting the aggregate itself.[2] The aggregation of forecasts of disaggregate inflation components is also receiving a lot of attention at central banks in the Eurosystem.[3] Similarly, for short-term inflation forecasting staff at the Federal Reserve Board focusses on estimating and forecasting disaggregate price categories.[4] The theoretical literature shows that aggregating component forecasts is at least as accurate as directly forecasting the aggregate if the data generating process is known. It does lower the mean squared prediction error (MSPE) except under certain conditions. 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.[5] Since in practice the data generating process is not known and given previous theoretical results, so far it remains largely an empirical question whether aggregating forecasts of disaggregates improves forecast accuracy of the aggregate of interest. The results in Hubrich (2005), for example, indicate that aggregating forecasts by component does not necessarily help to forecast year-on-year Eurozone inflation twelve months ahead.

In this paper we compare an alternative use of disaggregate information to forecast the aggregate variable of interest with methods considered in previous literature. We include disaggregate variables in the model for the aggregate, as we suggested in Hendry & Hubrich (2006). This is distinct from forecasting the disaggregate variables separately and aggregating those forecasts. An alternative to including disaggregate variables in the aggregate model might be to combine or summarize the in-

[2]See e.g. Fair & Shiller (1990) for a related analysis for US GNP, Zellner & Tobias (2000) for industrialised countries' GDP growth and Marcellino, Stock & Watson (2003) for disaggregation across euro area countries and forecasting as well as Espasa, Senra & Albacete (2002) and Hubrich (2005) for forecasting euro area inflation.

[3]See e.g. Bruneau, De Bandt, Flageollet & Michaux (2007), Benalal, Diaz del Hoyo, Landau, Roma & Skudelny (2004), Moser, Rumler & Scharler (2007) and Reijer & Vlaar (2006).

[4]See e.g. Bernanke (2007).

[5]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), Granger (1987), Pesaran, Pierse & Kumar (1989), Garderen, Lee & Pesaran (2000), Giacomini & Granger (2004); see also Granger (1990) for a survey on aggregation of time-series variables and Lütkepohl (2006) for a more recent review on forecasting aggregated processes by VARMA models.

formation contained in the disaggregate variables first and then include the resulting variables in the aggregate. Such a method of forecasting an aggregate by combining disaggregate information will entail a dimension reduction leading to reduced estimation uncertainty and, therefore, reduced mean squared forecast error. A factor model can be used, for example, which we include in our empirical analysis. The aim of the paper is to compare forecasting an aggregate by, first, combining disaggregate forecasts (considered in most previous related literature) with, second, including and combining disaggregate information in the aggregate model (our proposed competitor) and, third, only using lagged aggregate information (i.e. a univariate model for the aggregate).

Our analysis includes analytical derivations, simulation results and an empirical application to US inflation before and after the Great Moderation. It extends previous literature in a number of directions outlined in detail in the remainder of the introduction, also providing conditions under which a unique ranking between the different methods to forecast an aggregate is possible. The analysis is highly relevant for policy makers at central banks and policy observers interested in inflation forecasting since disaggregate inflation rates across sectors and regions are closely monitored and often used to forecast aggregate inflation. Many other applications of our results are possible including forecasting other macroeconomic aggregates such as GDP growth, monetary aggregates or trade, since our assumptions in a large part of the analysis are fairly general.

We illustrate the three methods for forecasting the aggregate that are compared in this paper using a simple autoregressive model, restricting the lag order to one for expositional purposes:

I) Forecasting the aggregate by lagged aggregates: $\hat{\pi}^{agg}_{T+1} = \hat{\gamma}_0 + \hat{\gamma}_1 \pi^{agg}_T$, i.e. employing an autoregressive model;

II) Forecasting disaggregates by lagged disaggregates, then aggregating those forecasts (as suggested in the previous literature on aggregation and forecasting): $\hat{\pi}^{agg}_{T+1} = \omega_1 \hat{\pi}^{(1)}_{T+1} + \omega_2 \hat{\pi}^{(2)}_{T+1}$, where the component forecasts $\hat{\pi}^{(1)}_{T+1}$ and $\hat{\pi}^{(2)}_{T+1}$ are aggregated (in this case a weighted sum);

III) Forecasting the aggregate based on lagged aggregates and lagged disaggregates (our proposal): $\hat{\pi}^{agg}_{T+1} = \hat{\delta}_0 + \hat{\delta}_1 \pi^{agg}_T + \hat{\beta}_1 \pi^{(1)}_T$, where the disaggregate variable $\pi^{(1)}_T$ is included in the aggregate model in addition to lags of the aggregate. Alternatively, the disaggregate $\hat{\pi}^{(2)}_T$ could be included. To avoid perfect collinearity, both disaggregates could only be included in the case of changing weights.

In contrast to predictability as a property in population, as analysed in the context of using disaggregate information for forecasting an aggregate in Hendry & Hubrich (2006), here we are interested in 'forecastability' referring to the improvement in forecast accuracy related to the sample information. Potential misspecification of the forecast model due to model specification, including the choice of variables included in the general information set and the choice of the functional form and model selection, estimation uncertainty as well as data measurement errors and structural breaks over the forecast horizon will affect the accuracy of the resulting forecast and help to explain why the population results on predictability are often not confirmed in empirical applications (see also Hendry (2004) and Clements & Hendry (2006)).

First, our proposal of combining disaggregate information by including all or a selected number of disaggregate variables in the aggregate model is investigating the predictability content of disaggregates for the aggregate from a new perspective. Most previous literature has instead focussed on combining disaggregate forecasts rather than disaggregate information. Combining disaggregate information gives rise to a classical model selection problem, while in previous work on combining disaggregate forecasts all disaggregates are included in the model and model selection is restricted to selecting the VAR order. Another alternative of combining disaggregate information that we consider in the empirical application is to use factors estimated from disaggregate components to forecast the aggregate.

Second, we present new analytical results for the forecast accuracy comparison of different uses of disaggregate information to forecast the aggregate. We include the additional method of forecasting the aggregate into the comparison of relative forecast accuracy, that is including disaggregate variables (or a variable combining disaggregate information to achieve dimension reduction) in the aggregate model. From the comparison of the forecast error decompositions of the different methods to forecast an aggregate we draw important conclusions regarding their relative forecast accuracy. We investigate the effect of misspecification and estimation uncertainty on the relative forecast accuracy of the different methods to forecast an aggregate.

Instabilities have generally been found important for the forecast accuracy of different forecasting methods, see e.g. Stock & Watson (1996), Clements & Hendry (1998, 1999, 2006) and Clark & McCracken (2006). Therefore, we extend the previous literature on contemporaneous aggregation and forecasting by allowing for a data generating process (DGP) with a break in the parameters and time-varying weights. Another extension of previous literature is that we consider the effect of unmodeled exogenous variables on the relative forecast accuracy of the different methods to forecast an aggregate.

Third, we present conditions under which a ranking between different approaches to forecast an aggregate is possible, i.e. the model is correctly specified and the DGP is assumed constant and we assume that we know the mean of the aggregate process or it is relatively constant.[6] This leads to reduced estimation uncertainty for the aggregate model and therefore lower mean square forecast error.

[6] For instance, when forecasting aggregate inflation in a country where the central bank has adopted an inflation targeting policy we might assume the target to be the mean of aggregate inflation.

Fourth, we analyze the effect of estimation uncertainty and misspecification and, in addition, the stochastic structure of the disaggregates and the interdependencies of the components on the relative forecast accuracy of the different approaches to forecast the aggregate by Monte Carlo simulations.

Fifth, we investigate whether our theoretical predictions can explain our empirical findings and analyze the relative forecast accuracy of combining disaggregate information versus disaggregate forecasts or just using past aggregate information to forecast aggregate US inflation.[7] [8]

The paper is organised as follows. In Section 2 we present new analytical results on the effect of changing coefficients, mis-specification and estimation uncertainty on the relative forecast accuracy of different approaches to forecast the aggregate. We also consider changing weights and present conditions for a unique ranking of the methods to forecast an aggregate. Section 3 presents Monte Carlo evidence for small samples. In Section 4, we investigate in a pseudo out-of-sample experiment for US inflation whether adding lagged values of the sub-indices of an aggregate improves the accuracy of forecasts of that aggregate relative to forecasting the aggregate only using lagged aggregate information, or aggregating forecasts of those sub-indices. Section 5 concludes.

2 Combining disaggregate forecasts versus combining disaggregate information

In this section we present new analytical results on a comparison of forecast errors made by first combining disaggregate forecasts, second, combining disaggregate information and third, only using past aggregate information when forecasting the aggregate is the objective. Let $\mathbf{y}_t$ denote the vector of n disaggregate price changes with elements $y_{i,t}$. We illustrate the analysis using as the DGP for the disaggregates an $\mathsf{I}(0)$ VAR with constant parameters in-sample extended by k unmodeled $\mathsf{I}(0)$ determinants $\mathbf{z}_{t-1}$:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\Phi}\mathbf{z}_{t-1} + \boldsymbol{\epsilon}_t \text{ for } t = 1, \ldots T \tag{1}$$

where $\boldsymbol{\epsilon}_t \sim \mathsf{ID}\left[\mathbf{0}, \boldsymbol{\Omega}\right]$ (ID: identically distributed), but consider a break at the forecast origin T with:

$$\mathbf{y}_{T+h} = \boldsymbol{\mu}^* + \boldsymbol{\Gamma}^*\mathbf{y}_{T+h-1} + \boldsymbol{\Phi}^*\mathbf{z}_{T+h-1} + \boldsymbol{\epsilon}_{T+h} \text{ for } h = 1, \ldots H \tag{2}$$

although the process stays $\mathsf{I}(0)$. Such a putative DGP reflects the prevalence of forecast failure in economics by its changing parameters, and allows an influence for both macroeconomic and policy

[7] For empirical findings on the effect of contemporaneous aggregation on forecast accuracy for the euro area, see e.g. Hubrich (2005).

[8] Results discussed in the text that are not presented explicitly in tables or graphs are available from the authors upon request.

variables. Let $y_t^a = \boldsymbol{\omega}_t' \mathbf{y}_t$ be the aggregate price index with weights $\boldsymbol{\omega}_t$. The aim of the following two sections is to compare the forecast errors from combining forecasts of the disaggregates with directly forecasting the aggregate from its past.

In Hendry & Hubrich (2006) we present analytical results on predictability of aggregates using disaggregates, focussing on predictability as a property in population. In the following analytical derivations, we allow the model and the DGP to differ and the parameters to be estimated. We extend previous literature by allowing for a structural break at the forecast origin[9] and allow for the inclusion of k unmodeled variables. Furthermore, we present conditions that allow a ranking of the different approaches to forecast an aggregate. Section 2.1 derives a forecast error decomposition for combining disaggregate forecasts, while Section 2.2 derives a comparable forecast error decomposition for forecasting the aggregate directly from its past. Section 2.3 compares the forecast error decompositions of the two methods to forecast an aggregate, draws five important conclusions. It also establishes conditions under which a unique ranking of the different methods is possible. This Section also discusses the implications of changing aggregation weights. Section 2.4 compares including disaggregate variables in the aggregate model with the methods to forecast an aggregate that are the main focus of previous literature and are analysed in sections 2.1 to 2.3.

2.1 Combining disaggregate forecasts: Some new analytical results

We first construct a taxonomy of the errors of the aggregated disaggregate forecasts using an estimated version of (1) when the forecast period is determined by (2). We present analytical results on the effect of parameter changes, misspecification and estimation uncertainty on the forecast error.[10] This analysis follows from the VAR taxonomy in Clements & Hendry (1998) extended to allow for the unmodeled variables, but we only consider 1-step forecasts when $(\mathbf{y}_T : \mathbf{z}_T)$ are known (forecast-origin uncertainty, such as implied by measurement errors) and multi-step ahead forecasts would add further terms, which we drop for readibility). Section 2.2 provides the corresponding taxonomy for the forecast errors of forecasting the aggregate directly from its past. In both cases, in-sample changes in the weights $\boldsymbol{\omega}_t$ make the analysis intractable. Therefore, we assume constant weights here, but will refer to the implications of changing weights later in Section 2.3.1. The weights are assumed to be positive and to lie in the interval [-1,1]. Five important new implications follow from our derivations

[9]We abstract from forecast origin uncertainties such as measurement errors, since they would complicate the analytics. Also, for our empirical application of forecasting aggregate inflation measurement errors are not very important.

[10]We do not consider data measurement error in these analytics since for our empirical application in this paper, i.e. for US inflation, it is not an important source of forecast error. However, in the empirical analysis in Section 4 we use factor models that deal with potential measurement errors.

and are discussed in Section 2.3.

First, for the aggregated disaggregate forecast taking expectations in (1) under stationarity[11] yields:

$$\mathsf{E}\left[\mathbf{y}_t\right] = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathsf{E}\left[\mathbf{y}_{t-1}\right] + \boldsymbol{\Phi}\mathsf{E}\left[\mathbf{z}_{t-1}\right] = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\phi}_y + \boldsymbol{\Phi}\boldsymbol{\phi}_z = \boldsymbol{\phi}_y,$$

which we refer to as the long-run mean.[12]

Therefore, $\boldsymbol{\phi}_y = \left(\mathbf{I}_n - \boldsymbol{\Gamma}\right)^{-1}\left(\boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\phi}_z\right)$ and $\boldsymbol{\phi}_z = \left(\mathbf{I}_n - \boldsymbol{\Phi}\right)^{-1}\left(\boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{\phi}_y\right)$ and hence:

$$\mathbf{y}_t - \mathsf{E}\left[\mathbf{y}_t\right] = \mathbf{y}_t - \boldsymbol{\phi}_y = \boldsymbol{\Gamma}\left(\mathbf{y}_{t-1} - \boldsymbol{\phi}_y\right) + \boldsymbol{\Phi}\left(\mathbf{z}_{t-1} - \boldsymbol{\phi}_z\right) + \boldsymbol{\epsilon}_t. \tag{3}$$

(3) represents the deviation of the disaggregates from their long-run mean, that will facilitate the decomposition of the forecast errors below and isolate terms that only affect the bias. Given the break after the forecast origin T in (2), however, the unconditional expectation of the forecast of the aggregate is:

$$\mathsf{E}\left[\mathbf{y}_{T+1}\right] = \boldsymbol{\mu}^* + \boldsymbol{\Gamma}^*\mathsf{E}\left[\mathbf{y}_T\right] + \boldsymbol{\Phi}^*\mathsf{E}\left[\mathbf{z}_T\right] = \boldsymbol{\mu}^* + \boldsymbol{\Gamma}^*\boldsymbol{\phi}_y + \boldsymbol{\Phi}^*\boldsymbol{\phi}_z$$

letting $\boldsymbol{\phi}_y^* = \left(\mathbf{I}_n - \boldsymbol{\Gamma}^*\right)^{-1}\left(\boldsymbol{\mu}^* + \boldsymbol{\Phi}^*\boldsymbol{\phi}_z^*\right)$ and $\boldsymbol{\phi}_z = \left(\mathbf{I}_n - \boldsymbol{\Phi}^*\right)^{-1}\left(\boldsymbol{\mu}^* + \boldsymbol{\Phi}^*\boldsymbol{\phi}_y\right)$, then the deviation of the disaggregate future values from their mean is given by (4):

$$\mathbf{y}_{T+1} - \mathsf{E}\left[\mathbf{y}_{T+1}\right] = \mathbf{y}_{T+1} - \boldsymbol{\phi}_y^* = \boldsymbol{\Gamma}^*\left(\mathbf{y}_T - \boldsymbol{\phi}_y^*\right) + \boldsymbol{\Phi}^*\left(\mathbf{z}_T - \boldsymbol{\phi}_z^*\right) + \boldsymbol{\epsilon}_{T+1}. \tag{4}$$

Next, from (3) the forecast vector of the disaggregate model is:

$$\widehat{\mathbf{y}}_{T+1|T} = \widehat{\boldsymbol{\phi}}_y + \widehat{\boldsymbol{\Gamma}}\left(\mathbf{y}_T - \widehat{\boldsymbol{\phi}}_y\right) + \widehat{\boldsymbol{\Phi}}\left(\mathbf{z}_T - \widehat{\boldsymbol{\phi}}_z\right) \tag{5}$$

where from $T+1$ onwards, the forecast errors $\widehat{\boldsymbol{\epsilon}}_{T+1|T} = \mathbf{y}_{T+1} - \widehat{\mathbf{y}}_{T+1|T}$ are:

$$\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T} = [\boldsymbol{\omega}'\boldsymbol{\phi}_y^* - \boldsymbol{\omega}'\widehat{\boldsymbol{\phi}}_y] + [\boldsymbol{\omega}'\boldsymbol{\Gamma}^*\left(\mathbf{y}_T - \boldsymbol{\phi}_y^*\right) - \boldsymbol{\omega}'\widehat{\boldsymbol{\Gamma}}\left(\mathbf{y}_T - \widehat{\boldsymbol{\phi}}_y\right)] + [\boldsymbol{\omega}'\boldsymbol{\Phi}^*\left(\mathbf{z}_T - \boldsymbol{\phi}_z^*\right) - \boldsymbol{\omega}'\widehat{\boldsymbol{\Phi}}\left(\mathbf{z}_T - \widehat{\boldsymbol{\phi}}_z\right)] + \boldsymbol{\omega}'\boldsymbol{\epsilon}_{T+1}. \tag{6}$$

Assuming the relevant moments exist, as is likely in the present context, let $\mathsf{E}[\widehat{\boldsymbol{\Gamma}}] = \boldsymbol{\Gamma}_e$ and $\mathsf{E}[\widehat{\boldsymbol{\phi}}_y] = \boldsymbol{\phi}_{y,e}$ etc.; furthermore $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}_e + \boldsymbol{\Xi}_{\boldsymbol{\Gamma}}$, $\widehat{\boldsymbol{\phi}}_y = \boldsymbol{\phi}_{y,e} + \boldsymbol{\xi}_{\phi}$, where $\boldsymbol{\Xi}$ and $\boldsymbol{\xi}$ represent the sampling variation of $\hat{\boldsymbol{\Gamma}}$ around $\boldsymbol{\Gamma}_e$ and $\hat{\boldsymbol{\phi}}_y$ around $\boldsymbol{\phi}_{y,e}$. Similar notation holds for the parameters of the unmodeled variables z_t ($\widehat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}_e + \boldsymbol{\Xi}_{\boldsymbol{\Phi}}$ and $\widehat{\boldsymbol{\phi}}_z = \boldsymbol{\phi}_{z,e} + \boldsymbol{\xi}_z$ etc.).

The forecast-error taxonomy now follows by decomposing each term in (6) into its components, i.e., the parameter shift, parameter mis-specification due to the small sample and estimation uncertainty of the parameters, as, for example, in the first term regarding the mean:

$$\boldsymbol{\omega}'\boldsymbol{\phi}_y^* - \boldsymbol{\omega}'\widehat{\boldsymbol{\phi}}_y = \boldsymbol{\omega}'\left(\boldsymbol{\phi}_y^* - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}'\left(\boldsymbol{\phi}_y - \boldsymbol{\phi}_{y,e}\right) + \boldsymbol{\omega}'\left(\boldsymbol{\phi}_{y,e} - \widehat{\boldsymbol{\phi}}_y\right), \tag{7}$$

[11] If the DGP is integrated, it must be transformed to a stationary representation.

[12] The expectation or long-run mean is referred to as ”equilibrium mean” in the work by Clements & Hendry (1998, 2006) . It is the value to which the process converges in the absence of further shocks. Nevertheless, the long-run mean might shift after a structural break.

where the objective is to isolate terms with zero and non-zero means respectively, to investigate the effects on the forecast bias and forecast variance.

The second term refers to the endogenous variables and can be decomposed as

$$\begin{aligned}\boldsymbol{\omega}'\boldsymbol{\Gamma}^*\left(\mathbf{y}_T-\phi_y^*\right)-\boldsymbol{\omega}'\widehat{\boldsymbol{\Gamma}}\left(\mathbf{y}_T-\widehat{\phi}_y\right) \;=\;& -\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}^*-\boldsymbol{\Gamma}\right)\left(\phi_y^*-\phi_y\right)-\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_e\right)\left(\phi_y^*-\phi_y\right)\\ & -\boldsymbol{\omega}'\boldsymbol{\Gamma}_e\left(\phi_y^*-\phi_y\right)\\ & +\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}^*-\boldsymbol{\Gamma}\right)\left(\mathbf{y}_T-\phi_y\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_e\right)\left(\mathbf{y}_T-\phi_y\right)\\ & -\boldsymbol{\omega}'\boldsymbol{\Gamma}_e\left(\phi_y-\phi_{y,e}\right)\\ & -\boldsymbol{\omega}'(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)\left(\mathbf{y}_T-\phi_y\right)-\boldsymbol{\omega}'(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)\left(\phi_y-\phi_{y,e}\right)\\ & +\boldsymbol{\omega}'\boldsymbol{\Gamma}_e(\widehat{\phi}_{\mathbf{y}}-\phi_{\mathbf{y},\mathbf{e}})+\boldsymbol{\omega}'(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)(\widehat{\phi}_{\mathbf{y}}-\phi_{\mathbf{y},\mathbf{e}})\end{aligned} \tag{8}$$

Finally, the third term refers to the exogenous variables, decomposed as

$$\begin{aligned}\boldsymbol{\omega}'\boldsymbol{\Phi}^*\left(\mathbf{z}_T-\phi_z^*\right)-\boldsymbol{\omega}'\widehat{\boldsymbol{\Phi}}\left(\mathbf{z}_T-\widehat{\phi}_z\right) \;=\;& -\boldsymbol{\omega}'\left(\boldsymbol{\Phi}^*-\boldsymbol{\Phi}\right)\left(\phi_z^*-\phi_z\right)-\boldsymbol{\omega}'\left(\boldsymbol{\Phi}-\boldsymbol{\Phi}_e\right)\left(\phi_z^*-\phi_z\right)\\ & -\boldsymbol{\omega}'\boldsymbol{\Phi}_e\left(\phi_z^*-\phi_z\right)\\ & +\boldsymbol{\omega}'\left(\boldsymbol{\Phi}^*-\boldsymbol{\Phi}\right)\left(\mathbf{z}_T-\phi_z\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Phi}-\boldsymbol{\Phi}_e\right)\left(\mathbf{z}_T-\phi_z\right)\\ & -\boldsymbol{\omega}'\boldsymbol{\Phi}_e\left(\phi_z-\phi_{z,e}\right)\\ & -\boldsymbol{\omega}'(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)\left(\mathbf{z}_T-\phi_z\right)-\boldsymbol{\omega}'(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)\left(\phi_z-\phi_{z,e}\right)\\ & +\boldsymbol{\omega}'\boldsymbol{\Phi}_e(\widehat{\phi}_{\mathbf{z}}-\phi_{\mathbf{z},\mathbf{e}})+\boldsymbol{\omega}'(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)(\widehat{\phi}_{\mathbf{z}}-\phi_{\mathbf{z},\mathbf{e}})\end{aligned} \tag{9}$$

Collecting terms from the decompositions in (7), (8) and (9) above and retaining the second-order terms of the form $(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)=\boldsymbol{\Xi}_{\boldsymbol{\Gamma}}, (\widehat{\phi}_{\mathbf{y}}-\phi_{\mathbf{y},\mathbf{e}})=\boldsymbol{\xi}_\phi$ etc. for completeness, yields:

Aggregated disaggregate forecast-error decomposition

$$\begin{aligned}&\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T}\simeq && \\ &\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}^*\right)\left(\phi_y^*-\phi_y\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}^*\left(\phi_z^*-\phi_z\right) && (ia)\ \textbf{long-run mean change}\\ &+\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}^*-\boldsymbol{\Gamma}\right)\left(\mathbf{y}_T-\phi_y\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Phi}^*-\boldsymbol{\Phi}\right)\left(\mathbf{z}_T-\phi_z\right) && (ib)\ \text{slope change}\\ &+\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}_e\right)\left(\phi_y-\phi_{y,e}\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}_e\left(\phi_z-\phi_{z,e}\right) && (iia)\ \textbf{long-run mean misspecification}\\ &+\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_e\right)\left(\mathbf{y}_T-\phi_y\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Phi}-\boldsymbol{\Phi}_e\right)\left(\mathbf{z}_T-\phi_z\right) && (iib)\ \text{slope mis-specification}\\ &+\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}_e\right)\left(\phi_{y,e}-\widehat{\phi}_y\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}_e(\phi_{z,e}-\widehat{\phi}_z) && (iiia)\ \text{long-run mean estimation}\\ &-\boldsymbol{\omega}'(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)\left(\mathbf{y}_T-\phi_{y,e}\right)-\boldsymbol{\omega}'(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)\left(\mathbf{z}_T-\phi_{z,e}\right) && (iiib)\ \text{slope estimation}\\ &+\boldsymbol{\omega}'(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)(\widehat{\phi}_y-\phi_{y,e})+\boldsymbol{\omega}'(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)(\widehat{\phi}_z-\phi_{z,e}) && (iv)\ \text{covariance interaction}\\ &+\boldsymbol{\omega}'\boldsymbol{\epsilon}_{T+1} && (v)\ \text{innovation error.}\end{aligned} \tag{10}$$

This forecast error decomposition allows separating terms that only affect the bias of the forecast. It also facilitates the analysis of the effect of structural change, mis-specification and estimation uncertainty, since certain terms vanish once no structural change or a correctly specified model is assumed. Terms with non-zero means affect the bias of the forecast and are shown in bold (see also Section 2.1.1). Equilibrium-mean mis-specification (*iia*) will only occur if there have been unmodeled in-sample mean shifts. These could be to a large extent avoided by careful in-sample modeling. Consequently, the mean change (*ia*) is the dominant source of the forecast bias. Changes in, and mis-specification of, the slope parameters in (*ib*), (*iib*) have zero-mean effects, also contributing to the forecast error variance. Parameter estimation uncertainty matters most when the DGP is constant and the model is well specified, but even so, its impact depends on the unknown values of the DGP parameters: see e.g., Calzolari (1987) and Clements & Hendry (1998) for derivations. The covariance interaction between the mean and the slope parameter estimation is represented in (*iv*). Interest here centers on forecasting the derived aggregate $y^a_{T+1} = \boldsymbol{\omega}'_{T+1}\mathbf{y}_{T+1}$ by $\widehat{y}^a_{T+1|T} = \widehat{\boldsymbol{\omega}}'_{T+1}\widehat{\mathbf{y}}_{T+1|T}$ so $(\boldsymbol{\omega}_{T+1}-\widehat{\boldsymbol{\omega}}_{T+1})$ is an additional source of error.

2.1.1 Expected values

Unconditional expectation: Implications for forecast bias Taking unconditional expectations of (10) results in all terms except for (*ia*) and (*iia*) being zero. Here we are interested in the effects on the bias of the aggregated disaggregate forecast. We will come back to the effects of estimation uncertainty on the variance in Section 2.1.2. The terms (*ia*) and (*iia*), that need not be zero unconditionally, are denoted by bold letters in (10). Then unconditionally,

$$\mathsf{E}\left[\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T}\right] = \boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}^*\right)\left(\boldsymbol{\phi}^*_y-\boldsymbol{\phi}_y\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}^*\left(\boldsymbol{\phi}^*_z-\boldsymbol{\phi}_z\right)+\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}_e\right)\left(\boldsymbol{\phi}_y-\boldsymbol{\phi}_{y,e}\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}_e\left(\boldsymbol{\phi}_z-\boldsymbol{\phi}_{z,e}\right).$$

Therefore, the aggregated disaggregate forecasts are unbiased if no mean shift and no mean misspecification occurs. The notation $\widehat{\epsilon}_{T+1|T}$ indicates a forecast error of a forecast made at time T.

Conditional expectation Taking the conditional expected value of the aggregated disaggregate forecast error in (10), interpreting $\boldsymbol{\phi}_{y,e}$ etc. as the mean estimates, i.e., ignoring the dependence of the estimates on the last observation, yields:

$$\begin{aligned}
&\mathsf{E}\left[\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T}\mid\mathbf{y}_T,\mathbf{z}_T\right] = && \\
&\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}^*\right)\left(\boldsymbol{\phi}^*_y-\boldsymbol{\phi}_y\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}^*\left(\boldsymbol{\phi}^*_z-\boldsymbol{\phi}_z\right) && (ia)\text{ long-run mean change}\\
&+\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}^*-\boldsymbol{\Gamma}_e\right)\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Phi}^*-\boldsymbol{\Phi}_e\right)\left(\mathbf{z}_T-\boldsymbol{\phi}_z\right) && (ib)\text{ slope change}\\
&+\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}_e\right)\left(\boldsymbol{\phi}_y-\boldsymbol{\phi}_{y,e}\right)-\boldsymbol{\omega}'\boldsymbol{\Phi}_e\left(\boldsymbol{\phi}_z-\boldsymbol{\phi}_{z,e}\right) && (iia)\text{ long-run mean misspecification}\\
&+\boldsymbol{\omega}'\left(\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_e\right)\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right)+\boldsymbol{\omega}'\left(\boldsymbol{\Phi}-\boldsymbol{\Phi}_e\right)\left(\mathbf{z}_T-\boldsymbol{\phi}_z\right) && (iib)\text{ slope mis-specification}\\
&+\boldsymbol{\omega}'\mathsf{E}\left[(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}_e)\left(\widehat{\boldsymbol{\phi}}_y-\boldsymbol{\phi}_{y,e}\right)\right]+\boldsymbol{\omega}'\mathsf{E}\left[(\widehat{\boldsymbol{\Phi}}-\boldsymbol{\Phi}_e)\left(\widehat{\boldsymbol{\phi}}_z-\boldsymbol{\phi}_{z,e}\right)\right] && (iv)\text{ covariance interaction}
\end{aligned}$$

where the last two terms, i.e. the covariance interaction, are likely to be small as the specification is relatively orthogonal being in terms of the long-run mean and deviations therefrom, and anyway are $\mathsf{O}_p\left(T^{-1}\right)$. Consequently, the conditional expectation will be affected by slope changes and slope misspecification in addition to mean change and misspecification, unlike the unconditional expectation. This implies that the bias of the optimal forecast is affected by the slope changes and slope misspecfication over the forecast period, while the first moment of the DGP is only affected by mean change and mean misspecification.

2.1.2 Forecast errors for constant parameter DGP and correctly specified model

Now we assume a constant-parameter DGP (i.e. $\mathbf{\Gamma}^* = \mathbf{\Gamma}$ and $\phi^* = \phi$) and correct specification (i.e. $\mathbf{\Gamma}_e = \mathbf{\Gamma}$ and $\phi_{y,e} = \phi_y$), with no unmodeled variables (i.e. any terms referring to unmodelled variables z_t in equation (10) is equal to zero) and the correct weights. In that case the error from combining disaggregate forecasts to forecast the aggregate is only due to estimation uncertainty, using term (iiia) and (iiib) in (10), taking into account the above assumptions:

$$\boldsymbol{\omega}'_{T+1}\widehat{\boldsymbol{\epsilon}}_{T+1|T} \simeq \boldsymbol{\omega}'_{T+1}\left(\mathbf{I}_n - \mathbf{\Gamma}\right)\left(\boldsymbol{\phi}_y - \widehat{\boldsymbol{\phi}}_y\right) + \boldsymbol{\omega}'_{T+1}\left[\left(\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}\right)\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)'\right] + \boldsymbol{\omega}'_{T+1}\boldsymbol{\epsilon}_{T+1}. \tag{11}$$

The unconditional and conditional expectations are zero, i.e., unbiased forecasts result under the above assumptions. The conditional forecast error variance is given by:

$$\begin{aligned}\mathsf{V}\left[\boldsymbol{\omega}'_{T+1}\widehat{\boldsymbol{\epsilon}}_{T+1|T} \mid \mathbf{y}_T\right] \;&\simeq\; \boldsymbol{\omega}'_{T+1}\mathbf{\Omega}\boldsymbol{\omega}_{T+1}\left[1 + T^{-1}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)'\mathbf{Q}^{-1}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)\right] \\ &\quad + T^{-1}\boldsymbol{\omega}'_{T+1}\left(\mathbf{I}_n - \mathbf{\Gamma}\right)\mathbf{\Omega}\left(\mathbf{I}_n - \mathbf{\Gamma}\right)'\boldsymbol{\omega}_{T+1}\end{aligned} \tag{12}$$

where, because of the assumption of correct specification, $\mathsf{V}\left[\widehat{\boldsymbol{\phi}}_y\right] \simeq T^{-1}\mathbf{\Omega}$, $\mathsf{V}\left[\widehat{\mathbf{\Gamma}}^{\nu}\right] \simeq T^{-1}\mathbf{\Omega} \otimes \mathbf{Q}^{-1}$ when $\mathsf{E}[(\mathbf{y}_t - \boldsymbol{\phi}_y)(\mathbf{y}_t - \boldsymbol{\phi}_y)'] = \mathbf{Q}$ and $\mathbf{Q} = \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}' + \mathbf{\Omega}$ is the standardized sample second-moment matrix (about means) of the disaggregates $\mathbf{y}_t$, and the covariances between the mean and the slope parameter estimation are essentially zero due to the decomposition in terms of deviations from the long-run mean. We now derive the forecast error when the aggregate is directly forecast from its past and compare it with (10) and (11) from combining disaggregate forecasts.

2.2 Forecasting the aggregate directly by its past

We consider forecasting the aggregate directly by its past, also allowing for exogenous variables z_t. First, pre-multiply (1) by $\boldsymbol{\omega}'_t$ to derive the aggregate relation:

$$\begin{aligned} y^a_t \;&=\; \boldsymbol{\omega}'_t\boldsymbol{\phi}_y + \boldsymbol{\omega}'_t\mathbf{\Gamma}\left(\mathbf{y}_{t-1} - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}'_t\mathbf{\Phi}\left(\mathbf{z}_{t-1} - \boldsymbol{\phi}_z\right) + \boldsymbol{\omega}'_t\boldsymbol{\epsilon}_t \\ &=\; \tau + \kappa\left(y^a_{t-1} - \overline{y^a}\right) + \nu_t \end{aligned} \tag{13}$$

say, where, in the second line, (τ, κ) orthogonalize $\left(1, y^a_{t-1} - \overline{y^a}\right)$ with respect to ν_t, and $\overline{y^a}$ is the mean of y^a. Hence:

$$\nu_t = \left(\boldsymbol{\omega}'_t \boldsymbol{\phi}_y - \tau\right) - \kappa \left(\boldsymbol{\omega}'_{t-1} \boldsymbol{\phi}_y - \overline{y^a}\right) + \left(\boldsymbol{\omega}'_t \boldsymbol{\Gamma} - \kappa \boldsymbol{\omega}'_{t-1}\right) \left(\mathbf{y}_{t-1} - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}'_t \boldsymbol{\Phi} \left(\mathbf{z}_{t-1} - \boldsymbol{\phi}_z\right) + \boldsymbol{\omega}'_t \boldsymbol{\epsilon}_t.$$

When the weights are relatively constant, $\boldsymbol{\omega}_t \simeq \boldsymbol{\omega}$, we have $\boldsymbol{\omega}' \boldsymbol{\phi}_y \simeq \overline{y}$ and hence $\tau = \boldsymbol{\omega}' \boldsymbol{\phi}_y$ so:

$$\nu_t \simeq \boldsymbol{\omega}' \left(\boldsymbol{\Gamma} - \kappa \mathbf{I}_n\right) \left(\mathbf{y}_{t-1} - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}' \boldsymbol{\Phi} \left(\mathbf{z}_{t-1} - \boldsymbol{\phi}_z\right) + \boldsymbol{\omega}' \boldsymbol{\epsilon}_t. \tag{14}$$

Non-constant weights add further terms to the error in (14), but also to (6).

A taxonomy of the sources of 1-step ahead forecast errors for y^a_{T+1} from a forecast origin at T can be derived from (14) which highlights the potential gains from adding disaggregates (or adding $\mathbf{z}_T$) to (13). As before, the forecast-period DGP is (2), so the innovation error $\boldsymbol{\omega}'_{T+1} \boldsymbol{\epsilon}_{T+1}$ has the same variance $\boldsymbol{\omega}'_{T+1} \boldsymbol{\Omega} \boldsymbol{\omega}_{T+1}$ as (11). Then:

$$y^a_{T+1} = \boldsymbol{\omega}'_{T+1} \boldsymbol{\phi}^*_y + \boldsymbol{\omega}'_{T+1} \boldsymbol{\Gamma}^* \left(\mathbf{y}_T - \boldsymbol{\phi}^*_y\right) + \boldsymbol{\omega}'_{T+1} \boldsymbol{\Phi}^* \left(\mathbf{z}_T - \boldsymbol{\phi}^*_z\right) + \boldsymbol{\omega}'_{T+1} \boldsymbol{\epsilon}_{T+1},$$

where:

$$\widetilde{y^a}_{T+1|T} = \widetilde{\tau} + \widetilde{\kappa} \left(y^a_T - \overline{y^a}\right).$$

Let $\boldsymbol{\omega}'_{T+1} \boldsymbol{\phi}^*_y = \tau_{T+1}$ and $\widetilde{\nu}_{T+1|T} = y^a_{T+1} - \widetilde{y}^a_{T+1|T}$ so using a similar notation as above:

$$\widetilde{\nu}_{T+1|T} = \tau_{T+1} - \tau + \boldsymbol{\omega}'_{T+1} \boldsymbol{\Gamma}^* \left(\mathbf{y}_T - \boldsymbol{\phi}^*_y\right) - \widetilde{\kappa} \left(y^a_T - \overline{y^a}\right) + \boldsymbol{\omega}'_{T+1} \boldsymbol{\Phi}^* \left(\mathbf{z}_T - \boldsymbol{\phi}^*_z\right) + \boldsymbol{\omega}'_{T+1} \boldsymbol{\epsilon}_{T+1}. \tag{15}$$

We again take the weights to be constant as an approximation to simplify (15) to a usable form so $\boldsymbol{\omega}' \boldsymbol{\phi}^*_y = \tau$:

Derived aggregate forecast-error decomposition

$$\begin{array}{ll}
\widetilde{\nu}_{T+1|T} \simeq & \\
\boldsymbol{\omega}' \left(\boldsymbol{\phi}^*_y - \boldsymbol{\phi}_y\right) - \boldsymbol{\omega}' \boldsymbol{\Gamma}^* \left(\boldsymbol{\phi}^*_y - \boldsymbol{\phi}_y\right) - \boldsymbol{\omega}' \boldsymbol{\Phi}^* \left(\boldsymbol{\phi}^*_z - \boldsymbol{\phi}_z\right) & (Ia)\ \textbf{long-run mean change} \\
+\boldsymbol{\omega}' \left(\boldsymbol{\Gamma}^* - \boldsymbol{\Gamma}\right) \left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}' \left(\boldsymbol{\Phi}^* - \boldsymbol{\Phi}\right) \left(\mathbf{z}_T - \boldsymbol{\phi}_z\right) & (Ib)\ \text{slope change} \\
+ \left(\tau - \tau_e\right) & (IIa)\ \textbf{long-run mean mis-specification} \\
+\boldsymbol{\omega}' \left(\boldsymbol{\Gamma} - \kappa_e \mathbf{I}_n\right) \left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) + \boldsymbol{\omega}' \boldsymbol{\Phi} \left(\mathbf{z}_T - \boldsymbol{\phi}_z\right) & (IIb)\ \text{slope mis-specification} \\
+ \left(\tau_e - \widetilde{\tau}\right) & (IIIa)\ \text{long-run mean estimation} \\
+ \left(\kappa_e - \widetilde{\kappa}\right) \boldsymbol{\omega}' \left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) & (IIIb)\ \text{slope estimation} \\
+\boldsymbol{\omega}' \boldsymbol{\epsilon}_{T+1} & (IV)\ \text{innovation error.}
\end{array} \tag{16}$$

As previously, the covariance interaction term is likely to be small due to the specification in terms of the long-run mean and its deviation therefrom. Thus, we neglect it here.

Terms in bold letters denote terms that need not be zero under unconditional expectations. Those terms would be zero only if no shift in the mean occured over the forecast period and a well specified model was used.

The conditional expectation will retain the terms of the mean and slope change as well as mean and slope misspecification, whereas the conditional variance depends on the slope misspecification and the two terms due to estimation uncertainty.

2.3 Comparing Aggregate versus Aggregated Disaggregate Forecast Errors: Conclusions and Conditions for unique ranking

2.3.1 Conclusions from forecast error comparison

Five important conclusions follow from comparing forecast errors from combining disaggregate forecasts in equation (10) with forecast errors from forecasting the aggregate directly (see equation (16)):

1. First (*Ia*) is identical to (*ia*). This implies that forecast-origin location shifts affect both methods of forecasting the aggregate in the same way. This is an important result to be noted since it is in contrast to previous literature on forecast combination if no disaggregate variables are involved. In that literature it is found that combining different forecasts for the same variable that are differentially biased might improve forecast accuracy in MSFE terms (see e.g. Clements & Hendry (2004)). Our analytical results show that this results does not carry over to combining disaggregate forecasts to forecast an aggregate.

2. Second, comparing (*Ib*) and (*ib*) shows that slope changes at the forecast origin do not affect the relative forecast accuracy of the different forecasting methods of the aggregate, either. Therefore, there are no gains or losses from aggregation in the presence of breaks at the forecast origin.

3. Third, the innovation error effects in (*IV*) and (*iv*) are also identical, irrespective of the covariance structure of the errors.

4. Fourth, long-run mean mis-specification is unlikely in both taxonomies when the in-sample DGP is constant.

5. Thus, fifth we conclude that slope mis-specification and estimation uncertainty are the primary source of forecast error differences.

Those conclusions will remain true for small changes in the weights $\boldsymbol{\omega}_T$ over the forecast horizon. Larger changes of weights $(\boldsymbol{\omega}_{T+1}-\widehat{\boldsymbol{\omega}}_{T+1})$ are an additional source of error. We leave more detailed investigation of this issue for future research.

2.3.2 Conditions for dominant role of estimation uncertainty

Following the fifth conclusion that slope misspecification and esitmation uncertainty are the primary source of forecast accuracy differences, we compare (16) with (10) assuming a constant-parameter DGP and correctly-specified disaggregate model when $\boldsymbol{\Gamma}_e = \boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}^* = \boldsymbol{\Phi} = \mathbf{0}$, so that no unmodeled exogenous variables influence the aggregate. Then the differences in the forecast errors are:

$$\begin{array}{lll} \boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T} - \widetilde{\nu}_{T+1|T} & & \\ \simeq -\boldsymbol{\omega}'\left(\boldsymbol{\Gamma} - \kappa_e \mathbf{I}_n\right)\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) & & \text{slope misspecification} \\ & & \text{of aggregate model} \\ +\boldsymbol{\omega}'\left(\mathbf{I}_n - \boldsymbol{\Gamma}\right)\left(\boldsymbol{\phi}_y - \widehat{\boldsymbol{\phi}}_y\right) & -\left(\tau - \widetilde{\tau}\right) & \text{long-run mean estimation} \\ \underbrace{+\boldsymbol{\omega}'(\boldsymbol{\Gamma} - \widehat{\boldsymbol{\Gamma}})\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)}_{\text{aggregated disaggregate forecasts}} & \underbrace{-\left(\kappa_e - \widetilde{\kappa}\right)\boldsymbol{\omega}'\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)}_{\text{direct forecast of aggregate}} & \text{slope estimation} \end{array} \tag{17}$$

The trade-off here is between the cost of omission by only using lags of the aggregate and the cost of estimation uncertainty.

Conditions for small slope misspecification of aggregate model Using the eigenvector decomposition $\boldsymbol{\Gamma} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^{-1}$ for the first row in (17), leads to

$$\boldsymbol{\omega}' H \left(\boldsymbol{\Lambda} - \kappa_e \mathbf{I}_n\right) H^{-1}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right).$$

κ_e solves a weighted eigenvalue problem that minimizes slope mis-specification of the aggregate model. A further interesting implication is that if all eigenvalues in $\boldsymbol{\Lambda}$ are equal, then slope misspecification will have a small effect on relative forecast accuracy of forecasting the aggregate directly or combining disaggregate forecasts.

To illustrate this, take the special case where in addition to the assumptions above we assume zero mean of the aggregate $\boldsymbol{\phi}_y = \mathbf{0}$, when $\mathbf{y}_t = \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$ and $y_t^a = \boldsymbol{\omega}'\mathbf{y}_t$, so:

$$y_t^a = \boldsymbol{\omega}'\boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\omega}'\boldsymbol{\epsilon}_t = \kappa y_{t-1}^a + \nu_t$$

then we get:

$$\kappa_e = \mathsf{E}\left[\widehat{\kappa}\right] = \mathsf{E}\left[\frac{\sum y_t^a y_{t-1}^a}{\sum\left(y_{t-1}^a\right)^2}\right] \xrightarrow{P} \frac{\text{plim}_{T\to\infty}\sum y_t^a y_{t-1}^a}{\text{plim}_{T\to\infty}\sum\left(y_{t-1}^a\right)^2} = \frac{\mathsf{E}\left[y_t^a y_{t-1}^a\right]}{\mathsf{E}\left[\left(y_{t-1}^a\right)^2\right]}$$

under stationarity, and hence:

$$\kappa_e \simeq \frac{\mathsf{E}\left[\left(\boldsymbol{\omega}'\boldsymbol{\Gamma}\mathbf{y}_{t-1}+\boldsymbol{\omega}'\boldsymbol{\epsilon}_t\right)\mathbf{y}_{t-1}'\boldsymbol{\omega}\right]}{\mathsf{E}\left[\boldsymbol{\omega}'\mathbf{y}_{t-1}\mathbf{y}_{t-1}'\boldsymbol{\omega}\right]} = \frac{\boldsymbol{\omega}'\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\omega}}{\boldsymbol{\omega}'\mathbf{Q}\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}'\mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}\mathbf{H}^{-1}\boldsymbol{\omega}}{\boldsymbol{\omega}'\mathbf{H}\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}\mathbf{H}^{-1}\boldsymbol{\omega}} = \frac{\boldsymbol{\phi}'\boldsymbol{\Lambda}\mathbf{P}\boldsymbol{\phi}}{\boldsymbol{\phi}'\mathbf{P}\boldsymbol{\phi}} \tag{18}$$

for $\boldsymbol{\phi}=\mathbf{H}^{-1}\boldsymbol{\omega}$, $\mathbf{Q}=E[\mathbf{y}_t\mathbf{y}_t']$ and $\mathbf{P}=\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}$. (18) can be written as:

$$\boldsymbol{\phi}'\left(\kappa_e\mathbf{I}_n-\boldsymbol{\Lambda}\right)\mathbf{P}\boldsymbol{\phi}=\mathbf{0},$$

so κ_e solves a weighted eigenvalue problem, precisely the one that minimizes the slope mis-specification in the taxonomy.

A number of the simulation experiments in Lütkepohl (1984) correspond to the case of equal eigenvalues, in which setting, estimation uncertainty differences will dominate forecast accuracy comparisons. We will use such a DGP also in our small sample simulations in Section 3.

2.3.3 Ranking of different methods: The effect of estimation uncertainty

In the following analytical derivations we present conditions under which a unique ranking between forecasting the aggregate with its own past and aggregating the component forecasts is possible. We focus on the effect of estimation uncertainty, thereby extending Kohn (1982) who establishes conditions for forecasting the aggregate efficiently from its past if the DGP is known. We use the nomenclature of (11) for ease of comparison, i.e. assuming constant parameter DGP and no mis-specification with implication of an unbiased forecast.[13] Again we exploit a zero covariance of mean and slope estimation, and using $\nu_t \sim \mathsf{D}\left[0,\sigma_v^2\right]$ where:

$$\begin{aligned}\sigma_v^2 &= \boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}+T^{-1}\left(\boldsymbol{\Gamma}-\kappa\mathbf{I}_n\right)\mathbf{Q}\left(\boldsymbol{\Gamma}-\kappa\mathbf{I}_n\right)'\boldsymbol{\omega}\\ &= \boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}+T^{-1}\boldsymbol{\omega}'\mathbf{H}\left(\boldsymbol{\Lambda}-\kappa\mathbf{I}_n\right)\mathbf{H}^{-1}\mathbf{Q}\mathbf{H}\left(\boldsymbol{\Lambda}-\kappa\mathbf{I}_n\right)'\left(\mathbf{H}^{-1}\right)'\boldsymbol{\omega}\simeq\boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}.\end{aligned} \tag{19}$$

Then the forecast error variance of the aggregated disaggregate forecasts is:

$$\begin{aligned}\mathsf{V}\left[\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T}\mid\mathbf{y}_T\right] &\simeq \boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega} && \text{variance of innovation}\\ &+T^{-1}\boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right)'\mathbf{Q}^{-1}\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right) && \text{slope variance}\\ &+T^{-1}\boldsymbol{\omega}'\left(\mathbf{I}_n-\boldsymbol{\Gamma}\right)\boldsymbol{\Omega}\left(\mathbf{I}_n-\boldsymbol{\Gamma}\right)'\boldsymbol{\omega} && \text{intercept variance.}\end{aligned} \tag{20}$$

In comparison the forecast error variance of the direct forecast of aggregate is:

$$\begin{aligned}\mathsf{V}\left[\widetilde{\nu}_{T+1|T}\mid\mathbf{y}_T\right] &\simeq \boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega} && \text{variance of innovation}\\ &+T^{-1}\sigma_v^2\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right)'\boldsymbol{\omega}\left(\boldsymbol{\omega}'\mathbf{Q}\boldsymbol{\omega}\right)^{-1}\boldsymbol{\omega}'\left(\mathbf{y}_T-\boldsymbol{\phi}_y\right) && \text{slope variance}\\ &+T^{-1}\sigma_v^2 && \text{intercept variance}\end{aligned} \tag{21}$$

[13] $\boldsymbol{\omega}=\boldsymbol{\omega}_{T+1}$ could be introduced by a change of notation.

Clearly, the first terms are equal. The second are directly comparable using the approximation in the second line of (19), letting the positive-definite matrix $\mathbf{Q} = \mathbf{K}\mathbf{K}'$ and $\mathbf{K}'\boldsymbol{\omega} = \mathbf{k}$ so:

$$\begin{aligned}\mathbf{Q}^{-1} - \boldsymbol{\omega}\left(\boldsymbol{\omega}'\mathbf{Q}\boldsymbol{\omega}\right)^{-1}\boldsymbol{\omega}' &= \left(\mathbf{K}\mathbf{K}'\right)^{-1} - \boldsymbol{\omega}\left(\boldsymbol{\omega}'\mathbf{K}\mathbf{K}'\boldsymbol{\omega}\right)^{-1}\boldsymbol{\omega}' \\ &= \left(\mathbf{K}'\right)^{-1}\left\{\mathbf{I}_n - \mathbf{k}\left(\mathbf{k}'\mathbf{k}\right)^{-1}\mathbf{k}'\right\}\mathbf{K}^{-1}.\end{aligned}$$

Then the difference is:

$$\begin{aligned}& T^{-1}\boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)'\mathbf{Q}^{-1}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) - T^{-1}\sigma_v^2\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)'\boldsymbol{\omega}\left(\boldsymbol{\omega}'\mathbf{Q}\boldsymbol{\omega}\right)^{-1}\boldsymbol{\omega}'\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right) \\ \simeq\ & T^{-1}\boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}\left[\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)'\left(\mathbf{K}'\right)^{-1}\left\{\mathbf{I}_n - \mathbf{k}\left(\mathbf{k}'\mathbf{k}\right)^{-1}\mathbf{k}'\right\}\mathbf{K}^{-1}\left(\mathbf{y}_T - \boldsymbol{\phi}_y\right)\right]\end{aligned}$$

where $\{\cdot\}$ is positive semi-definite idempotent, so the whole term must be non-negative. Finally, the difference between the third terms is:

$$T^{-1}\boldsymbol{\omega}'\left(\mathbf{I}_n - \boldsymbol{\Gamma}\right)\boldsymbol{\Omega}\left(\mathbf{I}_n - \boldsymbol{\Gamma}\right)'\boldsymbol{\omega} - T^{-1}\boldsymbol{\omega}'\left(\boldssymbol{\Gamma} - \kappa\mathbf{I}_n\right)\mathbf{Q}\left(\boldsymbol{\Gamma} - \kappa\mathbf{I}_n\right)'\boldsymbol{\omega} - \boldsymbol{\omega}'\boldsymbol{\Omega}\boldsymbol{\omega}.$$

Even if the middle component vanished, this term must still be negative, so $\mathsf{V}\left[\boldsymbol{\omega}'\widehat{\boldsymbol{\epsilon}}_{T+1|T} \mid \mathbf{y}_T\right] - \mathsf{V}\left[\widetilde{\nu}_{T+1|T} \mid \mathbf{y}_T\right]$ cannot be signed uniquely.

However, if we assume the long-run mean to be known, we can rank the two approaches to forecasting the aggregate uniquely. By assuming that the long-run mean of the aggregate is known, we assume that there is no contribution from the estimation of the intercept, i.e. the third term of the variance drops from the aggregate model (see equation (21)), but not from the disaggregate model. Then forecasting the aggregate directly will generally be preferable because of reduced estimation uncertainty.

An alternative condition for a unique ranking is the mean of the aggregate exhibiting very little volatility leading to lower estimation uncertainty of the mean estimate than for the mean of some of the very volatile disaggregate components.

As an example we argue that US aggregate inflation had a long-run mean of about 3% since the mid-80s, less volatile than most of its disaggregate components.

However, in addition to the particular case where the long-run mean of the aggregate is assumed to be known, changes in disaggregate weights, changes in the regressor correlation structure and/or the parameters of the disaggregate models will favour the aggregate model (see Hendry & Hubrich (2006) for some analytical results and Section 4 for an empirical illustration).

2.4 Combining disaggregate information to forecast the aggregate: Variable selection or dimension reduction

An alternative to the two methods to forecast an aggregate considered in the previous sections, i.e. combining disaggregate forecasts or forecasting the aggregate by past aggregate information, is to combine disaggregate information. One way to combine disaggregate information is to include disaggregate variables in the aggregate model.

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

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{y}_{t-1} + \mathbf{e}_t \tag{22}$$

as the DGP for the disaggregates. We assume constant weights and let $y_t^a = \boldsymbol{\omega}'\mathbf{y}_t$ be the aggregate price index with weights $\boldsymbol{\omega}$. Then pre-multiplying (1) by $\boldsymbol{\omega}'$:

$$\begin{aligned} y_t^a &= \boldsymbol{\omega}'\boldsymbol{\mu} + \boldsymbol{\omega}'\boldsymbol{\Gamma}\mathbf{y}_{t-1} + \boldsymbol{\omega}'\mathbf{e}_t = \boldsymbol{\omega}'\boldsymbol{\mu} + \kappa\boldsymbol{\omega}'\mathbf{y}_{t-1} + \left(\boldsymbol{\omega}'\boldsymbol{\Gamma} - \kappa\boldsymbol{\omega}'\right)\mathbf{y}_{t-1} + \boldsymbol{\omega}'\mathbf{e}_t \\ &= \boldsymbol{\omega}'\boldsymbol{\mu} + \kappa y_{t-1}^a + \left(\boldsymbol{\phi} - \kappa\boldsymbol{\omega}\right)'\mathbf{y}_{t-1} + \nu_t \end{aligned} \tag{23}$$

where $\boldsymbol{\phi} = \boldsymbol{\Gamma}\boldsymbol{\omega}$ are the constant parameters of the disaggregates in the aggregate model. In (23) the aggregate y_t^a depends on lags of the aggregate, y_{t-1}^a, and the lagged disaggregates $\mathbf{y}_{t-1}$. Thus, even if the DGP is (1) at the level of the components, an aggregate model will be systematically improved by adding disaggregates only to the extent that $\boldsymbol{\phi} - \kappa\boldsymbol{\omega} = \boldsymbol{\pi}$, i.e. the disaggregates contribute substantively to the explanation. The additional role of disaggregate information over just including the lagged aggregate in the aggregate model (23) is represented by the extent to which $\phi \neq \kappa\omega$ for each variable i. Time-varying weights would induce time-variation in the parameters. We present some analytical results on this issue in Hendry & Hubrich (2006).

If all but one disaggregate variables are added to the aggregate model, the combined disaggregate model is recovered with coefficients $\boldsymbol{\omega}'\boldsymbol{\Gamma}$ assuming constant weights. Thus, the general forecast error taxonomy (10) applies, as do the comparative conclusions drawn at the end of the previous section. Even if the DGP is (1) at the level of the components, an aggregate model will be systematically improved by adding disaggregates only to the extent that the coefficient matrix of the disaggregates is constant and contributes substantively to the explanation.

As we have shown in Section 2, estimation uncertainty of the mean and the slope parameters is the main determinant of relative forecast accuracy of the different methods to forecast the aggregate besides slope misspecification and can prevent an improvement of forecast accuracy due to disaggregates. Selection of a subset of the most relevant disaggregates to add to the model might help improving forecast accuracy of the aggregate largely by reducing estimation uncertainty. The forecast accuracy improvement depends on the explanatory power of the disaggregates in an R^2 sense.

An alternative to including disaggregate variables in the aggregate model might be to combine or summarize the information contained in the disaggregate variables first and then include the resulting variables in the aggregate. Such a method to forecasting an aggregate by combining disaggregate information will entail a dimension reduction leading to reduced estimation uncertainty and, therefore, to reduced mean squared forecast error.

3 Monte Carlo Simulations

This Monte Carlo simulation experiment is designed to compare forecasts of an aggregate by combining disaggregate and/or aggregate information with those based on combining disaggregate forecasts in small samples when the orders and/or coefficients of the data generating processes (DGPs) are not known. Lütkepohl (1984, 1987) is - to our knowledge - the only author to present small sample simulations on the effect of contemporaneous aggregation in forecasting.[14] We extend his simulations by including different DGPs as well as presenting results including the direct forecast of the aggregate from a VAR model including both the aggregate and a disaggregate component. The design of the DGPs aims to mimic features of the disaggregates and aggregate that we analyze in the empirical application.

The 2-dimensional DGP has the following general structure:

$$\begin{pmatrix} 1+\gamma_{11}L & \gamma_{12}L \\ \gamma_{21}L & 1+\gamma_{22}L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \tag{24}$$

where L is the backshift operator. The ν_{1t} and ν_{2t} are independent N(0,1) random numbers, therefore $\Sigma_\nu = I$. The aggregate is $y_t^a = y_{1,t} + y_{2,t}$.

For DGP 1 the parameters of DGP (24) are $\gamma_{11} = \gamma_{22} = 0.5$ and $\gamma_{12} = \gamma_{21} = 0$:

$$\begin{pmatrix} 1+0.5L & 0 \\ 0 & 1+0.5L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \tag{25}$$

Given those parameter values for the disaggregate processes, we get $(1+0.5L)y_t^a = 2 + \nu_t$ with $\sigma_\nu^2 = 2$ for the aggregate process. Thus, in DGP 1 the disaggregates y_1, y_2 and the aggregate y^a all follow an AR(1) process. Note that the eigenvalues of this process are equal. In this case slope misspecification will have a small effect on the relative forecast accuracy (see Section 2.3.2). This is the first DGP used in Lütkepohl (1984).

[14]Giacomini & Granger (2004) present simulation results for space-time aggregation.

The first DGP is constructed in such a way that the direct forecast of the aggregate and aggregating the disaggregate forecasts yield the same mean square forecast error (MSFE) since the components of the disaggregate multivariate process are independent and have identical stochastic structure. In those cases where the true model is used for estimation, the MSFE differences result only from estimation uncertainty and not from model mis-specification. Therefore, we can isolate the effect of the estimation uncertainty.

We modify DGP 1 to get DGP 2, 3 and 4, and extend to the 3-dimensional DGP 5. All five processes are stationary. The simulations are run for sample sizes of $T = 20, 40, 100, 200, 400$. We will only comment on results for $T = 20$ for DGP 1 and 5 and for T=100 for all DGPs in the text. The simulation results are based on N=1000 repititions in most cases.[15] As in Lütkepohl (1984) we generate forecasts from independent samples. A possible extension of the Monte Carlo simulations we present would be to estimate the models based on a recursively extended sample or a rolling sample. Some first results based on recursive samples did not qualitatively change the results presented in the following.

For DGP 1 the aggregate process follows an AR(p) model in contrast to DGP 2, DGP 3 and DGP 4, where the DGP of the aggregate follows an ARMA(2,1). Therefore, for DGP 1 the direct AR forecast of the aggregate exhibits higher forecast accuracy relative to the other methods, in contrast to DGP 2, DGP 3 and DGP 4, because the AR(p) model is misspecified for the latter three processes.

To further preview the results: For DGP 3 and DGP 4 including disaggregate information in the aggregate VAR model or forecasting the disaggregates from a VAR and aggregating their forecasts improves forecast accuracy over the other methods. For DGP3 both the direct and indirect AR forecasts perform similarly, while for DGP4 forecasting the aggregate directly from its past is somewhat better than the indirect AR forecast.

The higher the number of observations the closer the RMSFE of the direct and indirect forecast of the aggregate, as to be expected since, given the construction of the DGP implying equal population MSE, a larger number of observations leads to a decline in estimation uncertainty as well as the effect of lag order selection and therefore higher forecast accuracy. It is of interest to note that DGP 1 has a factor structure where the factor is y^a with equal weights of the disaggregates In this process the factor explains 100% of the variability in the data.

In Table 1 and 2 results from Monte Carlo simulations for DGP 1 are presented for a fixed true lag length of $p = 1$ for the aggregate and its components. The true lag length is used to isolate the effect of the estimation uncertainty from the effect of order selection on forecast accuracy. The results are presented for $T = 20$ and $T = 100$ observations. The direct forecast of the aggregate

[15] $N = 10000$ repititions are used for DGP 1 and DGP 5 for T=20 and fixed lag length of one.

based on the AR model is best, the indirect forecast based on the AR model performs second best, while the forecasts with the VAR model are the worst. This outcome is to be expected since the VAR models do not take the restriction into account that the disaggregate components are independent. For $T = 20$ the number of repetitions was increased to $N = 10000$ to investigate the ranking under reduced sampling variability. It turns out that the direct and indirect AR forecasts lie outside the 2 standard error confidence intervals of the VAR^{agg,sub} RMSFE for a horizon of $h = 1$ taking sampling variability into account, i.e. +/- $\sigma * \sqrt(2/N)$.

Further simulations are carried out using the AIC for the selection of the order of the model, since a model selection criterion would be employed in practice when the DGP is not known. We first comment on the results for the 1-step horizon for different samples when the lag length is selected by AIC. Table 3 shows that for T=100 observations the ranking between the different methods to forecast the aggregate are the same as for a lag length of one for a 1-step ahead horizon.

Our simulation results in terms of ranking are comparable to Lütkepohl (1987, Table 5.2), who also finds that the direct AR forecast is best for $T = 100$. In the case of DGP 1 to 4, where the true DGP is 2-dimensional, the VAR including the aggregate and one disaggregate component performs identically to the VAR with both subcomponents since the aggregate is a linear combination of the two disaggregates (see also Section 2.4).

Investigating the RMSFE for all horizons between $h = 1$ and 12 showed that the differences for horizons larger than 3 were minor. This is in line with the results presented in Lütkepohl (1984, 1987), who only presents results for $h = 1$ and $h = 5$. At forecast horizon of $h = 12$ all forecasts are almost identical.

In DGP 2 the disaggregates follow a VAR(1) process where y_1 and y_2 do have a different stochastic structure, but are still independent. The parameters of DGP (24) take the values $\gamma_{11} = 0.7$, $\gamma_{22} = 0.3$ and $\gamma_{12} = \gamma_{21} = 0$. The aggregate y^a follows an ARMA(2,1) process, i.e. $(1+L+0.21L^2)y_t^a = 3+\nu_t$ with $\nu_t = (1 + 0.3L)v_{1,t} + (1 + 0.7L)v_{2,t}$ and variance $\sigma_\nu^2 = 2.105$.

Note that estimating an AR(p) model for DGP 2 implies estimating a misspecified model for the aggregate. Table 4 shows that for T=100 the indirect AR based forecast of the aggregate is more accurate than the VAR forecasts and the direct forecast of the aggregate for a forecast horizon of one step ahead.

At forecast horizon of $h = 12$ all forecasts are almost identical again because the forecasts for longer horizons converge to the long-run mean of the process.

Note that even though the disaggregate components follow a different stochastic structure, the aggregate has a conditional variance very close to the sum of conditional variances of the disaggregates. The simulation results show that there is very little loss in fit and forecast in this case by using a forecast model approximating the aggregate process even though the differences in the eigenvalues are large.

In DGP 3 the disaggregate components are not independent in contrast to the previous DGPs, even though their autoregressive structure is the same. The parameters of DGP (24) take the values $\gamma_{11} = \gamma_{22} = 0.5$, $\gamma_{12} = -0.7$ and $\gamma_{21} = 0$. The aggregate y^a follows an ARMA(2,1) process, i.e. $(1 + 0.5L)^2 y_t^a = 3.7 + \nu_t$ with $\nu_t = (1 + 0.5L)v_{1,t} + (1 + 1.2L)v_{2,t}$ and variance $\sigma_\nu^2 = 2.562$.

For DGP 3 (see Table 5) disaggregate information helps forecasting the aggregate, in contrast to DGP 1 and DGP 2. Both the VAR with two subcomponents and the VAR with the aggregate and one subcomponent do perform better than the direct AR forecast of the aggregate, while direct and indirect AR perform similarly.

DGP 4 differs from DGP 3 due to the mutual dependence of the disaggregates. In DGP 3 only component $y_{1,t}$ depended on $y_{2,t}$. For DGP 4 the parameters of general DGP (24) take the values $\gamma_{11} = -0.5$, $\gamma_{22} = -0.3$, $\gamma_{12} = 0.6$ and $\gamma_{21} = 0.4$:

$$\begin{pmatrix} 1 - 0.5L & 0.6L \\ 0.4L & 1 - 0.3L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \tag{26}$$

with $\mathbf{\Sigma} = \mathbf{I}$ and $y_t^a = y_{1,t} + y_{2,t}$ the aggregate y^a follows an ARMA(2,1) process, i.e. $(1 - 0.8L - 0.09L^2)y_t^a = 0.2 + \nu_t$ with $\nu_t = (1 - 0.7L)v_{1t} + (1 - 1.1L)v_{2t}$ and variance $\sigma_\nu^2 = 2.277$.

Table 6 shows that for DGP 4 the VAR forecasts are again most accurate and the direct AR forecast is second best. It should be noted that, even though the DGP is stationary, the two eigenvalues are substantially different. DGP 4 reflects properties of some of the components in the application to US inflation in Section 4 since it is presenting 2 components with similar ACF, as commodities and services inflation.

Finally, we construct a three-dimensional DGP 5 where the disaggregates are independent and have equal stochastic structure as in DGP1. Again, differences in forecast accuracy between the different methods will therefore be due to estimation uncertainty.

$$\begin{pmatrix} 1 + 0.5L & 0 & 0 \\ 0 & 1 + 0.5L & 0 \\ 0 & 0 & 1 + 0.5L \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{pmatrix} \tag{27}$$

with

$$\Sigma_\nu = I, \quad y_t^a = y_{1,t} + y_{2,t} + y_{3,t},$$
$$(1 + 0.5L)y_t^a = 3 + \nu_t, \quad \sigma_\nu^2 = 2$$

For this DGP we can investigate the effect of variable selection on the VAR forecast. Table 1 and 7 for T=20 and T=100, respectively, show clearly an improvement in forecast accuracy if only one of the two disaggregates is included in the VAR. In both cases, the true lag length of one was chosen for the aggregate and the disaggregate models. When we use an order selection criterion instead (Table 8) the results are mixed since forecast accuracy in that case will be influenced by the model selection procedure in addition to estimation uncertainty.

Overall, we find that including disaggregate variables in the aggregate model or forecasting disaggregates from a VAR and aggregating the forecasts helps forecasting the aggregate if the disaggregates follow different stochastic structures and are interdependent. The differences in forecast accuracy are less pronounced for higher horizons, since then the forecast converges to the unconditional mean. Note that the simulations assume a constant parameter DGP with constant weights.

In practice, the correlation between the disaggregates depends on the level of disaggregation considered. However, the application for US inflation in Section 4 shows that even for a fairly low level of disaggregation, with only a few disaggregate components, at least some of the components are highly correlated and follow a similar stochastic structure.

4 Forecasting aggregate US inflation

In this section, we analyze empirically the relative forecast accuracy of the three methods to forecast the aggregate investigated analytically in the previous sections. We aim to answer the following questions regarding forecasting aggregate US inflation: First, does combining disaggregate information in the aggregate model improve over the direct forecast of the aggregate only using past aggregate information? We consider 2 methods of combining disaggregate information: Including disaggregate variables in the aggregate model or, alternatively, including factors estimated from disaggregate components of the aggregate. The latter approach is chosen since we have seen in Section 5 for DGP 5 that variable selection might improve forecast accuracy. Since the issue of model selection in forecasting goes beyond the scope of the paper, we use factor models to reduce estimation uncertainty. Second, is combining disaggregate information and including it in the aggregate model better in terms of forecast accuracy than forecasting disaggregate variables and aggregating those forecasts? We consider 3 methods of combining disaggregate forecasts: Combining disaggregate forecasts from AR or from VAR models using the weights of the components as well as combining single predictor models for each of the disaggregates with equal weights.

We analyze those issues using a broad range of forecast models and different model selection procedures and considering two very different sample periods for US inflation (see e.g. Atkeson &

Ohanian (2001) and Stock & Watson (2007) for recent contributions to predictability changes in US inflation). We investigate whether changes in aggregate US inflation and its components over those different sample periods do affect whether disaggregate information helps forecasting the aggregate. The sensitivity of our results with respect to various transformations of the variables in a model is investigated. We compare iterative and direct multi-step ahead forecasts and the effect on forecast accuracy using recursive versus rolling samples. We relate the findings to our theoretical results. The remainder of the section is organised as follows: Section 4.1 describes the data. Section 4.2 describes the broad range of forecast methods and model selection procedures employed. It presents details on the transformations used for building the forecast models and for forecast evaluation, including a sensitivity analysis comparing the results from the simulated out-of-sample forecast experiment for the level and the change in inflation. Results for different set-ups of the experiment based on recursively expanding or rolling estimation sample and direct versus iterative multi-step ahead forecasts are also discussed. Section 4.3 provide further details on the out-of-sample forecast experiment including information on the weights used for combining disaggregate forecasts. Section 4.4 presents the results from the empirical analysis of forecasting aggregate US inflation using vector autoregressive (VAR) models, including a discussion of the role of changing aggregation weights and the correlation structure. Section 4.5 presents the results for forecasting the aggregate using factor models and single predictor models based on a disaggregate information set. Finally, Section 4.6 summarises the empirical results and relates them to other empirical results in the literature.

4.1 Data

The data employed in this study include all items US consumer price index (CPI) as well as its breakdown into four subcomponents: food (p^f), commodities less food and energy commodities (p^c), energy (p^e) and services less energy services prices (p^s) (Source: CPI-U for all Urban Consumers, Bureau of Labor Statistics). We employ monthly, seasonally adjusted data.[16] We consider a sample period for inflation from 1960(1) to 2004(12), where earlier data from 1959(1) onwards are used for the transformation of the price level. As observed by other authors before (see e.g. Stock & Watson (2007)), there has been a substantial change in the mean and the volatility of aggregate inflation between the two samples. Figure 1 depicts aggregate year-on-year US inflation for all items CPI and its four subcomponents. It shows that also the disaggregate components exhibit a substantial change in mean and volatility. Aggregate as well as component inflation all exhibit high and volatile inflation

[16] We employed seasonally adjusted data, except for CPI energy that does not exhibit a seasonal pattern. Seasonal adjustment by the BLS is based on X-12-ARIMA.

until the beginning or mid 80s and lower, more stable inflation rates afterwards.

In Table 9 we show the substantial reduction in the mean for the disaggregate component inflation from the first to the second sample. The mean inflation rate has been reduced from 3.8-5.9% to 1.4-3.9%, while the standard deviation is reduced from between 2.9-8.2% to ranges of 1.0-8.3%. Thus, also the standard deviation has been reduced substantially, except for energy prices.

In section 4.3 we present results of an out-of-sample experiment for two different forecast evaluation periods: 1970(1) - 1983(12) and 1984(1) - 2004(12). The date 1984 for splitting the sample coincides with estimates of the beginning of the great moderation and is in line with what is chosen in Stock & Watson (2007) and Atkeson & Ohanian (2001). We use the same split sample for comparability of our results to those studies in terms of aggregate inflation forecasts.

The out-of-sample forecast evaluation period includes therefore 14 and 11 years for forecast evaluation.[17] ADF unit root tests do not reject non-stationarity of aggregate inflation, while it is rejected for the change in inflation. However, for the disaggregates the rejection of non-stationarity depends on the component and the sample. While for commodity inflation non-stationarity is rejected for the sample 1960-1983, it is rejected for services inflation for the second sample 1984-2004, when including a constant and a trend and lags up to the highest significant lag. For energy and food inflation non-stationarity is rejected most of the time. Due to the mixed results for the different components and samples and due to the low power of ADF tests in small samples, we carry out the forecast accuracy comparisons for the level and the change in inflation. We present the results for the level of inflation. The results for the changes of inflation do not differ much qualitatively from the results for the level of inflation.

4.2 Forecast methods and model selection

We have employed various forecasting methods using different model selection procedures for both direct and indirect forecast methods, i.e., forecasting HICP inflation directly versus aggregating subcomponent forecasts.

The forecasting models include 1. a simple autoregressive (AR) models; 2. the random walk (RW) implemented as inflation in $T + h$ being the simple average of the month-on-month inflation rate from $T - 12$ to T, as used in Stock & Watson (2007) referring to Atkeson & Ohanian (2001); 3. a subcomponent vector autoregressive model (VAR^{sub}) to indirectly forecast the aggregate by aggre-

[17]Note that we evaluate the 1- and 12-month ahead forecasts on the basis of the same forecast origin. This implies that the forecast evaluation period for 12-month ahead forecasts starts in 1971(1) and 1985(1), respectively, while the evaluation periods for 1-month ahead forecasts are 1970(1)-1983(12) and 1984(1)-2003(12).

gating subcomponent forecasts; 4. a VAR including the aggregate and the disaggregate components, VAR^{agg,sub}; 5. an MA(1) (as used in Stock & Watson (2007)); 6. factor models. We investigate empirically for US inflation the analytical findings from previous sections that including component information in the aggregate forecast model improves the forecast of the aggregate in some situations, but not in others.

The pseudo out-of-sample forecast results we present and discuss in the following are based on a recursively expanding estimation window. Note that model selection and estimation is carried out for each recursively expanded sample. The models selected are based on the AIC criterion due to the overall favorable forecast accuracy for US inflation (see section on model selection below). The multi-step ahead forecasts from the AR, RW and VAR models are derived using an iterative procedure. The results presented are based on models formulated in first differences and forecast accuracy is evaluated based on year-on-year forecasts. The results we present are based on the variable transformation, model selection procedure and set-up of out-of-sample experiment that resulted in the most accurate forecasts. In addition to those results presented and discussed in the paper, a large range of forecast experiments has been carried out. We explain and comment on additional results that we obtained in the following, after presenting the forecast equations.[18]

The iterative multi-step ahead forecast is based on the following model:

$$\widehat{\pi}_{T+1} = \widehat{\alpha} + \widehat{\beta}\pi_T + e_{T+1}, \tag{28}$$

$$\widehat{\pi}_{T+h} = \widehat{\alpha} + \sum_{i=0}^{p} \widehat{\beta}_i^h \pi_{T-i} + \sum_{i=1}^{h} \widehat{e}_{T+i} \tag{29}$$

where inflation π is specified in first differences $(P_t - P_{t-1})/P_{t-1}$. The forecast evaluation is based on a transformation of the resulting forecasts to year-on-year inflation $(\widehat{P}_{T+h} - \widehat{P}_{T+h-12})/\widehat{P}_{T+h-12}$.

We also include results for factor models and single predictor models. The factors are estimated from disaggregate price information by principal components. The factors are assumed to be uncorrelated and the idiosyncratic term is assumed to have limited correlation across units and over time in order to satisfy the conditions in Stock & Watson (2002a, 2002b) . Under those assumptions the model is identified and the factors and loadings can be estimated. Note that other related literature has also focused on the approximate factor model and shown consistency of principal components estimators of the factor space, e.g. Bai (2003), Bai & Ng (2002) and Forni, Hallin, Lippi & Reichlin

[18] The results are available from the authors upon request.

(2000, 2005) . [19] [20]

We compute the direct h-step DFM forecasts as

$$\widehat{\pi}_{T+h}^{h} = \widehat{\alpha}_{(h)} + \sum_{i=1}^{p} \widehat{\phi}_{i(h)} \pi_{T-i} + \sum_{j=1}^{r} \widehat{\theta}_{j(h)} \widehat{F}_{T-j+1} + \widehat{\epsilon}_{T+h} \tag{30}$$

where $\widehat{\pi}_{T+h}^{h}$ denotes the forecast of the rate of inflation $\pi_t^h = (logP_t - logP_{t-h}) = h^{-1} \sum_{i=1}^{h-1} \pi_{t-i}$ and $\widehat{F}_t$ are the estimated factors. In addition, the direct forecasts based on a single predictor Z_t are derived from

$$\widehat{\pi}_{T+h}^{h} = \widehat{\alpha}_{(h)} + \sum_{i=1}^{p} \widehat{\phi}_{i(h)} \pi_{T-i} + \sum_{k=1}^{l} \widehat{\theta}_{k,(h)} Z_{T-k+1} + \widehat{\epsilon}_{T+h} \tag{31}$$

The forecast is evaluated based on π_{T+h}^{h}. We also consider the forecast combination of all single predictor models based on the respective disaggregate component with equal weights.

These results for the dynamic factor models are not directly comparable with the iterative multi-step ahead forecasts based on the VARs, except for a horizon of $h = 12$ months. This is because here direct multi-step ahead forecasts are carried out and forecast accuracy is evaluated for annualised inflation (instead of year on year inflation for each horizon as in the previous tables). We thereby match the specification used in Stock & Watson (1999).

Model selection Model selection procedures selecting the lag length in the various models employed above include the Schwarz (SIC) and the Akaike (AIC) criterion, respectively.[21] For a discussion of in-sample versus out-of-sample forecast model selection the reader is referred to Inoue & Kilian (2005). We find that the AIC based models generally perform better for US inflation, in particular for the AR(p) over all horizons and both samples. Note, however, that e.g. for the sample 1970-1983 the SIC selected AR models with higher forecast accuracy for higher horizons than the AIC.

Expanding versus rolling window First, we compared recursive estimation with an expanding estimation window to forecasts based on a rolling estimation sample where the window size is fixed

[19] For treatments of classical factor models when the cross-sectional dimension n is small, see e.g. Anderson (1984), Geweke (1977), Sargent & Sims (1977), Stock & Watson (n.d.); see Doz, Giannone & Reichlin (2006) on the relation of principal components estimators and quasi-maximum likelihood estimators.

[20] Note that a larger cross-section relative to T improves the asymptotic performance in the sense that consistency is achieved at a faster rate as in the case of a small cross-section (see Stock & Watson (1998)). Since we wanted to keep our information set comparable with the one underlying the forecast experiments with the VAR models, we did not increase the number of disaggregate variables.

[21] The maximum lag order was chosen to be 13 due to the monthly frequency of the data.

to 10 years. Overall, we find that the recursive scheme provides more accurate forecasts of aggregate inflation than the rolling scheme. This implies that the reduction in variance due to using more information for the estimation dominates over the bias due to breaks, i.e. estimation uncertainty is more important than breaks.

Direct versus iterative multi-step forecast Furthermore, we have investigated the relative forecast accuracy of the direct versus iterative procedure for forecasting multi-step ahead for the AR model. We find mixed results. The results depend on the sample. For 1970-1983 the direct multi-step ahead forecast is better, while for 1984-2004 the iterative multi-step ahead forecast provides slightly more accurate forecasts.

Month-on-month versus year-on-year transformation Finally, we carried out a comparison of different variable transformations used in the models. More specifically, we compared forecasts based on models specified in terms of month-on-month inflation and models specified in terms of year-on-year inflation. Note, however, that the evaluation of the relative forecast accuracy is in both cases carried out on the basis of year-on-year inflation, the variable of interest for central banks, since the ranking of the different forecast methods is not invariant to the selected transformations (see e.g. Clements & Hendry, 1998, pp 68). The relative forecast accuracy of these transformations important for forecasting inflation has, to our knowledge, not been investigated so far. We find that month-on-month inflation modeling is best overall if evaluated based on year-on-year inflation. This can be interpreted as an indication that the ”true” frequency of inflation can rather be considered monthly than annual, which might be in line with what is to be expected from an economic point of view. In that case, where the true DGP of inflation is in first differences of prices, the year-on-year transformation would imply moving average error terms. This implies the additional disadvantage of distorted inference in the model formulated in terms of year-on-year inflation.

Level of inflation versus change of inflation We repeated our forecast experiments for the level of inflation presented in the next section for the change in inflation. The results are not qualitatively different, in particular regarding the ranking of combining disaggregate information or combining disaggregate forecast or just using lags of aggregate inflation to forecast the aggregate.

4.3 The experiments

Simulated out-of-sample forecast experiments are carried out to evaluate the relative forecast accuracy of alternative methods to forecast aggregate US inflation combining information on its disaggregate components as opposed to combining the forecasts of 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 or rolling 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 squared forecast error (RMSFE).

Combination of disaggregate forecasts The combination of the disaggregate forecasts for the AR, RW, MA and VAR models is implemented by replicating the aggregation procedure employed by the BLS for the CPI disaggregate data. The data are aggregated in terms of levels taking into account the respective base year of the weights. Historical aggregation weights were provided to the authors by the BLS. Historical weights and their changes over time are depicted in Figure 10. For the aggregation of the forecasts the weights, that would be known to the forecaster in real time, are used. For the single predictor models forecasts are combined taking the simple average.

4.4 Combining disaggregate forecasts or disaggregate variables: AR and VAR models

Table 10 and 11 present the comparison of the forecast accuracy measured in terms of RMSFE of year-on-year (headline) US inflation of forecasting aggregate (all items) inflation using the different approaches analyzed and discussed throughout the paper. The benchmark for the comparison is the (direct) forecast of aggregate inflation based on the AR model (first entry in column labelled 'direct'), i.e. simply forecasting aggregate inflation from its past. This is compared to other methods of forecasting the aggregate directly, e.g. the random walk and the VAR including the aggregate and subcomponents.[22] Furthermore, we compare the 'direct' AR forecast with the 'indirect' forecast of aggregate inflation (column labelled 'indirect') , i.e. the aggregated forecasts of the sub-indices from both subcomponent VARs excluding or including the aggregate (VAR^{sub} and $VAR^{agg,sub}$). We also present results for the direct and indirect forecast of the aggregate using an MA(1).

$\Delta_{12}\widehat{p}^{agg}$ and $\Delta_{12}\widehat{p}_{sub}^{agg}$ indicate that the forecast is evaluated on the basis of year-on-year inflation. The models are, however, specified in terms of month-on-month inflation. Multi-step forecasts 6- and 12-months ahead presented in this section are derived in an iterative procedure. Values below one for the relative RMSFE indicate an improvement of the forecast based on the respective method over the direct AR forecast.

[22] Note that perfect collinearity between aggregate and components does not pose a problem due to annually changing weights in price indices.

Overall, we find that the direct forecast of the aggregate is more accurate than the indirect forecast of the aggregate. The direct aggregate forecast based on the $AR(p)$, RW and the $VAR^{agg,sub}$ appear to be similar for a forecast horizon of $h = 1$ for the forecast evaluation period 1970-1983, while for the period 1984-2004 they are broadly similar for all horizons. In contrast, for the earlier period for higher horizons the $AR(p)$ provides the most accurate forecasts. The MA(1) is worse than the $AR(p)$ in the first sample period and similar to it in the second sample. We find similar results for the first difference of inflation. Therefore, we do not find for monthly CPI inflation that the MA(1) for the level and the change in inflation outperforms the benchmark direct AR forecast. This finding in line with Stock & Watson (2007), who analyze four different price measures and find that quarterly CPI inflation is not well modeled by an IMA(1,1). They raise the issue of temporal aggregation in this context as a possible explanation. An improvement over the AR forecast for the one year ahead forecast matches our result for the first difference of year-on-year inflation, which indicates that the ranking of the MA versus the AR model very much depends on the transformation used.

When relating those empirical results to our earlier analytical findings, estimation uncertainty might be considered as important in driving those results. We did not use any variable selection procedure for the disaggregates to reduce the number of parameters to be estimated in the $VAR^{agg,sub}$. Such a strategy might improve forecast accuracy of using disaggregate information to forecast an aggregate. However, when we estimated many different bivariate and trivariate VARs with disaggregate components, we found little imporvement in terms of forecast accuracy over the $VAR^{agg,sub}$ and over the AR model. In any case, in the second sample after the great moderation estimation uncertainty might be less relevant for higher horizons due to the stationarity of the inflation process (see e.g. Stock (1996) and Chong & Hendry (1986)), explaining lower RMSFE for all methods for that period.

We also carried out a forecast error decomposition for the AR model and the random walk. The results are presented in Table 12 and 13. We find that the direct AR model has smaller bias and smaller forecast error variance for $h = 1$ in both samples. For the earlier high inflation period the AR model has a lower bias than the RW, while for the low inflation period the RW provides forecasts with lower bias. The indirect AR forecast has smaller variance for $h = 6$ and 12.

Overall, for the high inflation sample including the 1970s we find that the direct AR model dominates the indirect both in terms of bias and variance. For the latter sample 1984-2004 the bias effect of the indirect forecast dominates the effect of a small forecast error variance of the indirect forecast, so that the direct AR forecast is again most accurate. The RMSFE results indicate that generally the direct forecast is more accurate than the indirect forecast. However, the direct and the indirect RMSFE are much closer in terms of RMSFE in the more recent sample 1984-2004 than in the sample 1970-1983, since in a low inflation environment where inflation exhibits very little variability it is very difficult to improve over simple univariate forecasting models.

For illustration we present the graphs of the direct and the indirect (aggregated) forecast of the aggregate for both samples in Figure 2 to 5. Those graphs show that steep upturns or downturns of inflation are less well forecasted by aggregating component forecasts.

4.4.1 Changes in component weights and correlation structure

In Section 2 we have commented on the effect of changing weights on forecast accuracy of forecasting the aggregate. Furthermore, in Hendry & Hubrich (2006) we have analyzed analytically the effects of different types of changes influencing forecast accuracy of the aggregate model including disaggregate components. We now analyze two of those changes in the context of forecasting US inflation: a change in component weights and a change in collinearity of disaggregate regressors.

Figures 10 and 11 exhibit the weights of US CPI components in levels as well as the weights over the standard deviation of the components, respectively. There are large changes about every 10 years at the beginning and smaller changes every two years starting in 2001. Table 14 presents the ratio of the percentage change in weight and the standard deviation of month-on-month inflation of the respective CPI component. The relation to the standard deviation provides a measure of the relative size and importance of the weight change for the respective component. The largest changes in weights take place in 1986 where a change in CPI measurement has taken place, i.e. some components previously included in commodities are now included in the services index. Apart from that, in the 70s there are changes of up to -6.2% for CPI food and +3.2% for commodities, while in e.g. 2000 the relative weight changes span from -1.4% for food and +2.7% for commodities. Therefore, the changes in weights do influence the forecast accuracy. However, the infrequent changes in the 70s and 80s imply that for the first out-of-sample period 1970-1983 the weight changes affect the forecasts only for a year. For the second out-of-sample period 1984-2004 the changes in weights are, in contrast, more important due to size and frequency of the changes.

These changes in weights mean that the relevance of the changes of, say, food prices for the aggregate declines over the forecast evaluation period so that positive shocks to food prices do affect the aggregate less, whereas the positive shocks to services prices will affect the aggregate more in the future. [23]

Second, we analyze the change in the correlation structure between the aggregate and the components over the forecast evaluation period. Figure 6 presents the correlation of US all items CPI

[23] The indirect forecast of the aggregate by aggregating the component forecasts is also affected since the weights are used for aggregation.

with its components. The graph clearly shows that there are correlation changes between the estimation and the out-of-sample periods as well as within the out-of-sample periods for both samples considered. Similarly, the correlation between subcomponents are depicted in Figure 7 to 9, where the correlations with the least changes are shown in the respective graphs. In particular, the correlation between energy inflation and commodities as well as food inflation (Figure 7) change repeatedly and dramatically over both sample periods, although more before 1983.

The above effects favour an aggregate model, in particular for longer forecast horizons like 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 in forecast model selection is unknown a priori.

4.5 Disaggregate information in dynamic factor models

A different method of combining disaggregate information to forecast an aggregate is to include factors estimated from the disaggregate components in the aggregate model. This approach is used in the following and also compared to the benchmark AR model. The motivation for such an approach is that we found in the analytical investigation estimation uncertainty to be an important determinant of the relative forecast accuracy of the different methods to forecast an aggregate. By using factor models we reduce estimation uncertainty in comparison with the VAR with many parameters.

We employ factor models averaging away idiosyncratic variation in the disaggregate series. We include the estimated factors in the aggregate model. Little is known so far how the size and the composition of the data affect the factor estimates. Some results indicating that more data are not always better for factor analysis can be found in Boivin & Ng (2005). In this paper we are concerned with how factors from disaggregate information affect forecast accuracy of the aggregate economic variable. As discussed in Section 4.2, we estimate the factors by principal components. The models considered in this section are also more parsimonious than many VARs considered in the previous section and therefore the forecast accuracy is less affected by estimation uncertainty. Please note that these results are not directly comparable across all horizons with the previous table except for h=12, since here direct multi-step ahead forecasts are carried out and forecast accuracy is evaluated for annualised inflation in line with Stock & Watson (1999, 2007) (instead of year on year inflation as in the previous tables). We compute the direct h-step factor forecasts and single predictor forecasts. We also consider forecast combination of all single predictor models based on the respective disaggregate component with equal weights. The results for US year-on-year inflation are presented in Tables 15 and 16.

For the first sample, disaggregate information does help forecasting aggregate US inflation one and twelve months ahead. The improvements over the AR model are up to 6.5% in RMSFE terms (up to 12.5% in MSFE terms). For the second sample period the improvement using factor models is lower than for the first period. This is in line with what Stock & Watson (2007) find for including real variables in the inflation model. For the second period there is less variability in the aggregate to be explained by the disaggregates, while in the first sample period the disaggregates contain more information useful for forecasting the aggregate. Table 11 shows that for the second sample period the random walk provides slightly more accurate forecasts than the AR model for a horizon of 12 months (see also Atkeson & Ohanian (2001)), but not 1 month ahead or in the first sample period. However, comparing with the RMSFE ratios in Table 16 indicates that the factor model is improving somewhat more in terms of forecast accuracy over the AR model than the random walk.

4.6 Summary of Empirical Results

In Table 9 we showed the substantial reduction in the mean and the standard deviation of aggregate and disaggregate component inflation (except for energy inflation) from the first to the second sample. Therefore, it is not surprising that all the forecast results show more accurate forecasts for the second sample with lower RMSFE independent of the forecast method used, i.e. independent of the model used and also independent of whether a direct or an indirect method of forecasting the aggregate has been employed.

To summarize our empirical results, we find that overall the direct forecast of the aggregate is more accurate than the indirect. All forecasts are quite close for 1 month ahead prediction, although the direct forecast with the AR and the factor models using disaggregate information are doing better than the other models. Some single predictor models improve slightly over the AR model only in some situations, while the RW does for the 12 months ahead forecast in the second sample period. The factor model exhibits the highest improvement in forecast accuracy over the AR model, in particular in the first sample period. Therefore, the empirical results indicate that for forecasting inflation in those situations, where disaggregate information helps forecasting the aggregate, combining disaggregate information helps over combining disaggregate forecasts.

Regarding the question how to combine disaggregate information to improve forecasts, one possibility explored in this empirical application is using factor models. Another possibility might be to employ variable selection procedures to improve on the forecast of e.g. the $VAR^{sub,agg}$. This is beyond the scope of the paper. Bivariate and trivariate VARs based on disaggregate variables indicated little improvement over the direct AR forecast.

Other empirical literature on contemporaneous aggregation and forecasting As mentioned in the introduction, there is a growing empirical literature on contemporaneous aggregation and forecasting. Fair & Shiller (1990), for instance, analyze relative forecast accuracy for real US GNP growth and find that disaggregate information helps forecasting the aggregate. There are a number of studies on euro area countries or the euro area as a whole (see Introduction for citations). Those studies mostly find that information on disaggregate variables might help forecasting the aggregate, but only for particular models or model selection procedures or certain forecast horizons (see e.g. Hubrich (2005)). In a related study by Marcellino et al. (2003) the authors analyze empirically the issue of aggregation versus disaggregation in forecasting using disaggregate country information when forecasting inflation and real activity of the euro area. They find that pooling country-specific univariate autoregressions provides the most accurate forecasts in their application. The reasons for the difference to our results might be explained by the fact that country aggregation is considered rather than the aggregation of CPI component forecasts, and second that a particular sample period is considered for euro area data, as pointed out by the authors, that is the period of convergence before EMU 1982-1997. Substantial differences in political and economic developments made country-specific information more essential. On the other hand, similar to our results, the authors also find that the univariate forecast tend to outperform the multivariate forecast, and factor forecasts tend to outperform VARs.

5 Conclusions

In this paper, we present new analytical results on the relative forecast accuracy of combining disaggregate forecasts, i.e. first forecasting the disaggregate variables separately and then aggregating those forecasts, relative to forecasting an aggregate using only lagged aggregate information.

We compare these methods, that have been the focus of previous literature, with an alternative use of disaggregate information to forecast the aggregate variable of interest, that is to combine disaggregate information by including all or a selected number of disaggregate components in the aggregate model. We suggested this approach of combining disaggregate information to forecast an aggregate in Hendry & Hubrich (2006), where we investigate predictability in population in this context.

Our analysis in this paper includes new analytical derivations, small sample simulation results and an empirical application to US inflation before and after the Great Moderation. We choose inflation forecasting as an empirical application since this is highly relevant for policy makers at central banks and policy observers interested in inflation forecasting since disaggregate inflation rates across sectors and regions are closely monitored and often used to forecast aggregate inflation. Our results, however, are relevant for many other applications, for instance forecasting other macroeconomic aggregates such as output growth, monetary aggregates or trade.

In Hendry & Hubrich (2006) we show that a forecasting model including both aggregate and disaggregate variables in the predictor set should in population provide lower or at least equal prediction mean squared error relative to a model that includes only lags of the aggregate or relative to first predicting the disaggregates and then aggregating those predictions.

In contrast, in this paper we are investigating the improvement in forecast accuracy related to sample information. When the forecast model differs from the data generation process and / or the parameters of the model are not known, and have to be estimated from the data, the predictive value of disaggregate information can be off-set.

Therefore, we present new analytical results on the effects of mis-specification and estimation uncertainty on the relative forecast accuracy of the different approaches to forecast an aggregate. We extend previous results by allowing for a change in the parameters of the DGP and time-varying weights. From the comparison of the forecast error decomposition of the different methods to forecast an aggregate we draw five important conclusions regarding their relative forecast accuracy. First, we find that their relative forecast accuracy is not affected by forecast-origin location shifts. This is in contrast to the forecast combination literature that focuses on combining forecasts of the same variable where combination helps in the presence of mean shifts in opposite directions. Second, we find that there are no gains or losses from aggregation of component forecasts with respect to slope changes. Third, relative forecast accuracy is not affected by innovation errors, irrespective of the covariance structure. Fourth, we think that mean-misspecification is unlikely to be an important source of forecast error for both methods to forecast an aggregate when the in-sample DGP is constant or in-sample shifts are modeled. Fifth, we conclude that slope misspecification and estimation uncertainty are the primary sources of differences in forecast accuracy between the different methods. We also consider the role of changing weights as an additional source of forecast error.

Furthermore, we present conditions under which a ranking between different approaches to forecast an aggregate is possible, allowing for estimation uncertainty. More specifically, we show that forecasting the aggregate from past aggregate information provides the most accurate forecast under the assumption of a constant parameter DGP and well-specified models, if the mean of the aggregate is known. Not having to estimate the mean of the aggregate does reduce the estimation uncertainty and thereby the MSFE of the aggregate forecasting model. Alternatively, the mean of the aggregate might be relatively constant leading to lower estimation uncertainty of the mean estimate than for some of the more volatile disaggregate components.

We also analyze the effect of estimation uncertainty and mis-specification as well as the stochastic structure of the disaggregates and the interdependencies between the components on the relative forecast accuracy by Monte Carlo simulations. We consider a number of DGPs relevant for the issue of combining disaggregate variables or disaggregate forecasts to forecast the aggregate. We find that including disaggregate variables in the aggregate model helps forecasting the aggregate if the disaggregates follow different stochastic structures and the components are interdependent.

Finally, we investigate whether our theoretical predictions can explain our empirical findings by analysing the relative forecast accuracy of 1. combining disaggregate forecasts versus 2. combining disaggregate information or 3. just using past aggregate information to forecast aggregate US inflation. We find that changing weights in the price index and changing correlations between disaggregate prices undermine the performance of disaggregate-based models. Furthermore, combining disaggregate information tends to outperform combining disaggregate forecasts when forecasting aggregate inflation. In particular, we find that summarising the information in the disaggregates in estimated factors improves forecast accuracy of the aggregate. This seems to confirm our analytical finding that estimation uncertainty plays an important role for the relative forecast accuracy of the different approaches to forecast an aggregate, since by using factor models we reduce the number of the parameters to be estimated.

Including disaggregate information in the aggregate model might help for forecasting. Overall we find that estimation uncertainty is an important determinant of the relative forecast accuracy of the different methods to forecast an aggregate. Therefore, we recommend model selection procedures for the disaggregates to be included in the aggregate model or methods to combine disaggregate information.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edn, Wiley, New York.

Atkeson, A. & Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation?, *Federal Reserve Bank of Minneapolis Quarterly Review* **25**(1): 2–11.

Bai, J. (2003). Inference on factor models of large dimensions, *Econometrica* **71**(1): 135–172.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

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.

Bernanke, B. (2007). Inflation expectations and inflation forecasting, *Speech at the Monetary Economics Workshop of the NBER Summer Institute*.

Boivin, J. & Ng, S. (2005). Are more data always better for factor analysis?, *Journal of Econometrics, forthcoming*.

Bruneau, C., De Bandt, O., Flageollet, A. & Michaux, E. (2007). Forecasting inflation using economic indicators: The case of France, *Journal of Forecasting* **26**: 1–22.

Calzolari, G. (1987). Forecast variance in dynamic simulation of simultaneous euqations models, *Econometrica* **55**: 1473–1476.

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

Clark, T. E. & McCracken, M. W. (2006). Forecasting with small macroeconomic VARs in the presence of instabilities, *Research Working Papers 06-09*, The Federal Reserve Bank of Kansas City.

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

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

Clements, M. P. & Hendry, D. F. (2006). Forecasting with breaks, *in* G. Elliott, C. W. J. Granger & A. Timmermann (eds), *Handbook of Economic Forecasting, forthcoming*, Elsevier.

Doz, C., Giannone, D. & Reichlin, L. (2006). A quasi maximum likelihood approach for large approximate dynamic factor models, *ECB Working Paper 674*.

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.

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

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2005). The generalized factor model: One-sided estimation and forecasting, *Journal of the American Statistical Association, forthcoming*.

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

Geweke, J. F. (1977). The dynamic factor analysis of economic time series, *in* D. J. Aigner & A. S. Goldberger (eds), *Latent Variables in Socio-Economic Models*, North-Holland, Amsterdam.

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

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

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

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

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

Hendry, D. F. & Hubrich, K. (2006). Forecasting aggregates by disaggregates, *European Central Bank Working Paper 589*.

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

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

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. (2006). Forecasting with VARMA processes, *in* G. Elliott, C. W. J. Granger & A. Timmermann (eds), *Handbook of Economic Forecasting, forthcoming*, Elsevier.

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

Moser, G., Rumler, F. & Scharler, J. (2007). Forecasting Austrian inflation, *Economic Modelling, forthcoming*.

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.

Reijer, A. & Vlaar, P. (2006). Forecasting inflation: An art as well as a science!, *De Economist* **127**(1): 19–40.

Sargent, T. J. & Sims, C. A. (1977). Business cycle modeling without pretending to have too much a-priori theory, *in* C. Sims (ed.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

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. (1996). Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics* **14**(1): 11–30.

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

Stock, J. H. & Watson, M. W. (1999). Forecasting inflation, *Journal of Monetary Economics* **44**: 293–335.

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

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

Stock, J. H. & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast?, *Journal of Money, Credit and Banking, forthcoming*.

Stock, J. H. & Watson, M. W. (n.d.). A probability model of the coincident economic indicator.

Zellner, A. & Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance, *Journal of Forecasting* **19**: 457–469.

Table 1: **Actual RMSFE for DGP 1 and DGP 5, $h=1$, true lag order $p=1$**

	DGP 1		DGP 5	
	$\gamma_{11}=\gamma_{22}$=0.5 $\gamma_{12}=\gamma_{21}$=0		$\gamma_{11}=\gamma_{22}=\gamma_{33}$=0.5 γ_{ij}=0 for $i \neq j$	
sample size	$T=20$		$T=20$	
method	direct	indirect	direct	indirect
$\text{AR}_{(1)}$	1.489	1.496	1.828	1.826
$\text{VAR}^{sub}_{(1)}$		1.548		1.961
$\text{VAR}^{agg,sub(y1)}_{(1)}$	1.548		1.886	
$\text{VAR}^{agg,sub(y2)}_{(1)}$			1.892	
$\text{VAR}^{agg,sub(y1,y2)}_{(1)}$			1.961	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_1)}$: VAR with aggregate and subcomponent y1; true lag order $p=1$ for all models, $N=10000$

Table 2: **Actual RMSFE for DGP 1, $T=100$, true lag order $p=1$**

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(1)}$	1.406	1.409	1.629	1.631	1.656	1.656
$\text{VAR}^{sub}_{(1)}$		1.412		1.631		1.656
$\text{VAR}^{agg,sub(y1)}_{(1)}$	1.412		1.631		1.656	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_1)}$: VAR with aggregate and subcomponent y1; true lag order $p=1$ for all models, $N=1000$, DGP 1: $\gamma_{11}=\gamma_{22}=0.5$ and $\gamma_{12}=\gamma_{21}=0$

Table 3: **Actual RMSFE for DGP 1,** $T = 100$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(AIC)}$	1.408	1.422	1.634	1.643	1.656	1.655
$\text{VAR}^{sub}_{(AIC)}$		1.424		1.637		1.657
$\text{VAR}^{agg,sub(y_1)}_{(AIC)}$	1.424		1.637		1.657	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_1)}$: VAR with aggregate and subcomponent y_1; Subscripts indicate lag selection, AIC: Akaike criterion, $N = 1000$, DGP 1: $\gamma_{11} = \gamma_{22} = 0.5$ and $\gamma_{12} = \gamma_{21} = 0$

Table 4: **Actual RMSFE for DGP 2,** $T = 100$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(AIC)}$	1.442	1.423	1.729	1.736	1.770	1.771
$\text{VAR}^{sub}_{(AIC)}$		1.428		1.726		1.773
$\text{VAR}^{agg,sub(y_1)}_{(AIC)}$	1.428		1.726		1.773	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_1)}$: VAR with aggregate and subcomponent y_1; Subscripts indicate lag selection, AIC: Akaike criterion, DGP 2: $\gamma_{11} = 0.7$, $\gamma_{22} = 0.3$ and $\gamma_{12} = \gamma_{21} = 0$

Table 5: **Actual RMSFE for DGP 3,** $T = 100$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(AIC)}$	1.604	1.605	1.744	1.771	1.660	1.655
$\text{VAR}^{sub}_{(AIC)}$		1.417		1.731		1.661
$\text{VAR}^{agg,sub(y_1)}_{(AIC)}$	1.417		1.731		1.661	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_1)}$: VAR with aggregate and subcomponent y_1; Subscripts indicate lag selection, AIC: Akaike criterion, $N = 1000$, DGP 3: $\gamma_{11} = \gamma_{22} = 0.5$, $\gamma_{12} = -0.7$ and $\gamma_{21} = 0$

Table 6: **Actual RMSFE for DGP 4,** $T = 100$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(AIC)}$	1.524	1.696	1.547	1.641	1.560	1.614
$\text{VAR}^{sub}_{(AIC)}$		1.425		1.507		1.545
$\text{VAR}^{agg,sub(y_1)}_{(AIC)}$	1.425		1.507		1.545	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_i)}$: VAR with aggregate and subcomponent y_1. Subscripts indicate lag selection, AIC: Akaike criterion, $N = 1000$, DGP 4: $\gamma_{11} = -0.5$, $\gamma_{22} = -0.3$, $\gamma_{12} = 0.6$, $\gamma_{21} = 0.4$

Table 7: **Actual RMSFE for DGP 5,** $T = 100$**, true lag order** $p = 1$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(1)}$	1.753	1.752	2.033	2.035	2.059	2.059
$\text{VAR}^{sub}_{(1)}$		1.770		2.034		2.059
$\text{VAR}^{agg,sub(y_1)}_{(1)}$	1.768		2.033		2.059	
$\text{VAR}^{agg,sub(y_2)}_{(1)}$	1.764		2.034		2.059	
$\text{VAR}^{agg,sub(y_2,y_1)}_{(1)}$	1.770		2.034		2.059	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_i)}$: VAR with aggregate and subcomponent y_i; Subscripts indicate lag selection, $N = 1000$, DGP 5: $\gamma_{11} = \gamma_{22} = \gamma_{33}$=0.5, γ_{ij}=0 for $i \neq j$

Table 8: **Actual RMSFE for DGP 5,** $T = 100$

horizon	1		6		12	
method	direct	indirect	direct	indirect	direct	indirect
$\text{AR}_{(AIC)}$	1.770	1.766	2.038	2.039	2.062	2.061
$\text{VAR}^{sub}_{(AIC)}$		1.778		2.034		2.060
$\text{VAR}^{agg,sub(y_1)}_{(AIC)}$	1.783		2.039		2.061	
$\text{VAR}^{agg,sub(y_2)}_{(AIC)}$	1.768		2.039		2.061	
$\text{VAR}^{agg,sub(y_2,y_1)}_{(AIC)}$	1.778		2.034		2.060	

Note: Superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub(y_i)}$: VAR with aggregate and subcomponent y_i; Subscripts indicate lag selection, AIC: Akaike criterion, $N = 1000$, DGP 5: $\gamma_{11} = \gamma_{22} = \gamma_{33}$=0.5, γ_{ij}=0 for $i \neq j$

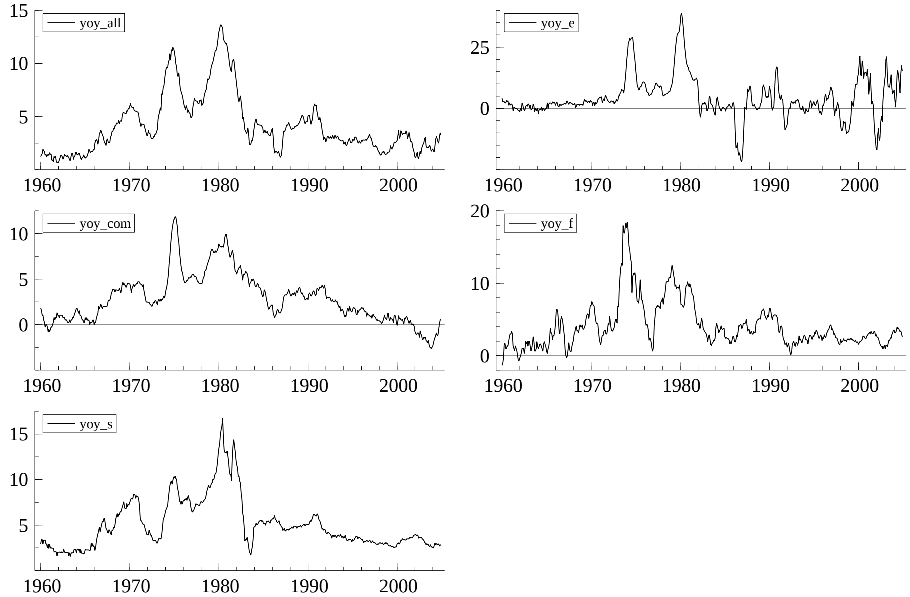

Figure 1: Year-on-year US CPI inflation rate, aggregate and subcomponents

Table 9: US, Descriptive Statistics, year-on-year CPI Inflation

1960-1983	all items	energy	commodities	food	services
Mean	4.86	5.91	3.80	4.75	5.81
Std Deviation	3.41	8.17	2.89	4.11	3.40
1984-2004	all items	energy	commodities	food	services
Mean	2.99	2.28	1.43	2.93	3.91
Std Deviation	1.06	8.26	1.65	1.26	0.99

Table 10: **Relative RMSFE, US year-on-year inflation (percentage points), 1970-1983**

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 $\Delta_{12}\widehat{p}^{agg}_{sub}$
$\text{AR}_{(AIC)}$	**0.294**	1.337	**1.358**	1.083	**2.985**	1.324
RW	1.031	1.378	1.053	1.048	1.045	1.061
$\text{MA}_{(1)}$	1.395	1.198	1.899	1.828	1.695	1.318
$\text{VAR}^{sub}_{(AIC)}$		1.450		1.241		1.429
$\text{VAR}^{agg,sub}_{(AIC)}$	1.071	1.468	1.129	1.225	1.254	1.437

Note: Actual RMSFE (non annualised) for $\text{AR}_{(AIC)}$ model in percentage points, for other models RMSFE relative to AR; recursive estimation samples 1960(1) to 1970(1),...,1983(12), maximum number of lags: $p = 13$; Subscripts indicate model selection procedure, AIC: Akaike criterion, superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub}$: VAR with aggregate and subcomponents; 'direct': direct forecast of the aggregate, 'indirect': aggregated subcomponent forecast

Table 11: **Relative RMSFE, US year-on-year inflation (percentage points), 1984-2004**

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 $\Delta_{12}\widehat{p}^{agg}_{sub}$
$\text{AR}_{(AIC)}$	**0.190**	1.528	**0.685**	1.024	**1.261**	1.021
RW	1.000	1.617	0.994	1.095	0.955	0.997
$\text{MA}_{(1)}$	1.037	1.508	1.129	1.134	1.116	1.021
$\text{VAR}^{sub}_{(AIC)}$		1.627		1.155		1.102
$\text{VAR}^{agg,sub}_{(AIC)}$	1.044	1.610	1.107	1.179	1.074	1.111

Note: Actual RMSFE (non annualised) for $\text{AR}_{(AIC)}$ model in percentage points, relative RMSFE for other models; recursive estimation samples 1960(1) to 1984(1),...,2004(12), maximum number of lags: $p = 13$; Subscripts indicate model selection procedure, AIC: Akaike criterion, superscripts indicate model, VAR^{sub}: VAR only including subcomponents, $\text{VAR}^{agg,sub}$: VAR with aggregate and subcomponents;'direct': direct forecast of the aggregate, 'indirect': aggregated subcomponent forecast

Table 12: MFE and Forecast Error Variance (FEV) for yoy inflation, US, 1970-1983

	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 $\Delta_{12}\widehat{p}^{agg}_{sub}$
MFE	AR$_{(AIC)}$	**-0.005**	0.092	**-0.023**	-0.424	**-0.125**	-1.403
	RW	0.013	0.179	0.099	0.279	0.209	0.391
FEV	AR$_{(AIC)}$	**0.090**	0.137	**1.871**	1.949	8.165	**8.161**
	RW	0.092	0.132	2.034	1.949	9.688	9.880

Note: MFE and FE Variance in percentage points, Recursive estimation samples 1959(01) to 1970(01),....1983(12), maximum number of lags allowed: p=13; Subscripts indicate model selection procedure, AIC: Akaike criterion, model: mom inflation

Table 13: MFE and Forecast Error Variance for yoy inflation, US, 1984-2004

	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 $\Delta_{12}\widehat{p}^{agg}_{sub}$
MFE	AR$_{(AIC)}$	**0.021**	-0.037	**0.206**	0.301	**0.588**	0.910
	RW	**0.006**	-0.068	0.038	**-0.031**	0.084	**0.012**
FEV	AR$_{(AIC)}$	**0.037**	0.082	0.482	**0.409**	1.440	**0.978**
	RW	**0.036**	0.090	0.462	0.561	1.445	1.579

Note: MFE and FE Variance in percentage points, Recursive estimation samples 1959(01) to 1984(01),....2004(12), maximum number of lags allowed: p=13; Subscripts indicate model selection procedure, AIC: Akaike criterion, model: mom inflation

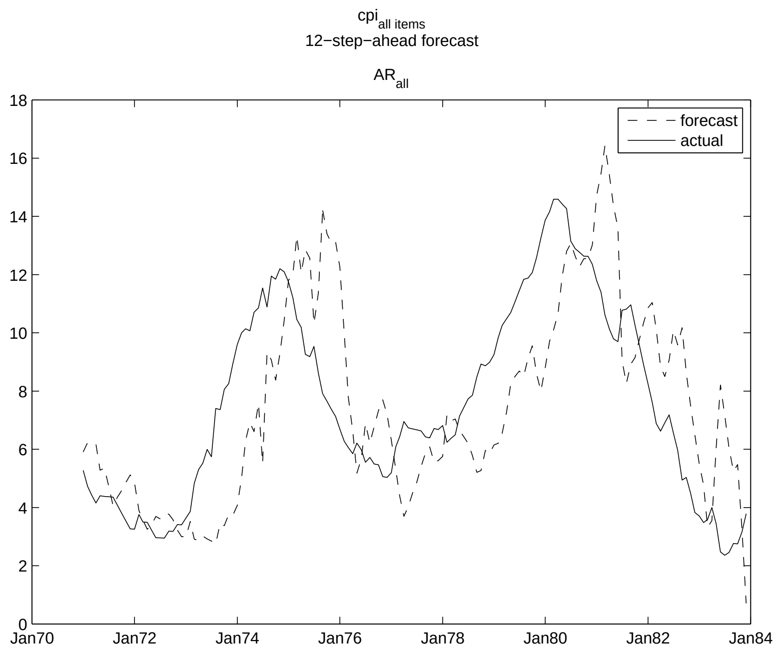

Figure 2: US year-on-year inflation rate and direct forecast of the aggregate in %, 12 months ahead, solid line: actual, 1970-1983

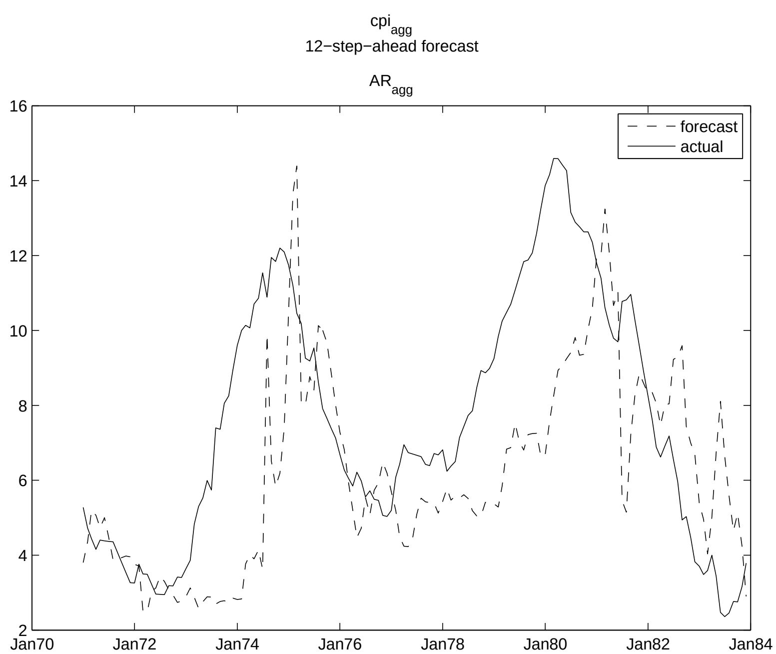

Figure 3: US year-on-year inflation rate and indirect forecast of the aggregate in %, 12 months ahead, solid line: actual, 1970-1983

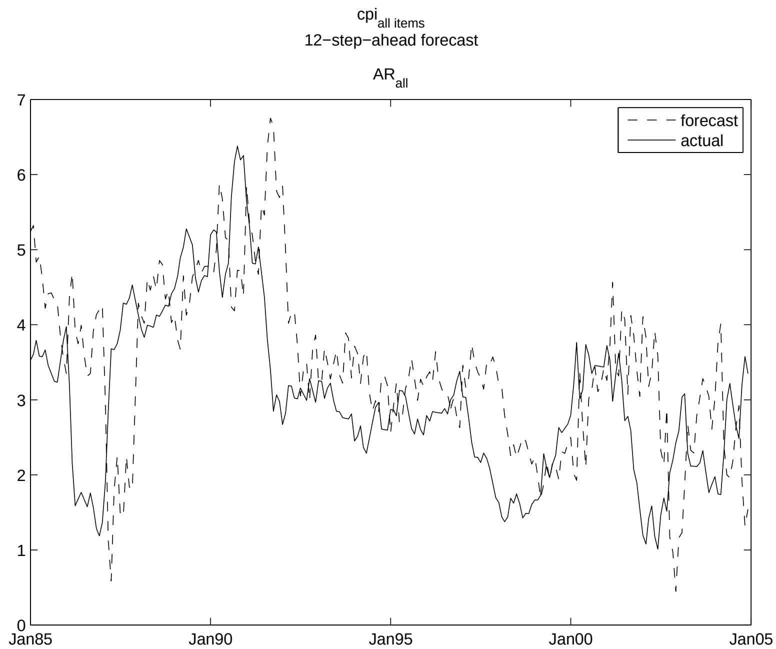

Figure 4: US year-on-year inflation rate and direct forecast of the aggregate in %, 12 months ahead, solid line: actual, 1984-2004

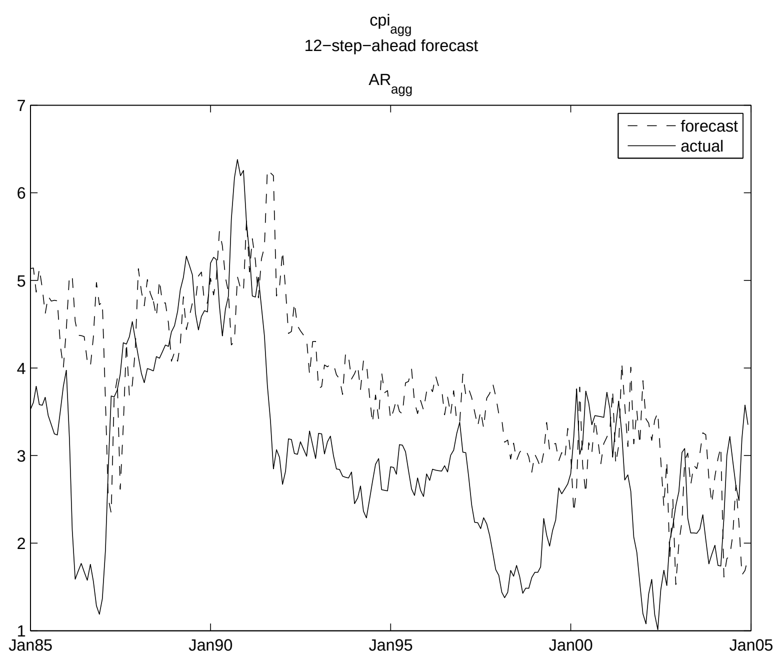

Figure 5: US year-on-year inflation rate and indirect forecast of the aggregate in %, 12 months ahead, solid line: actual, 1984-2004

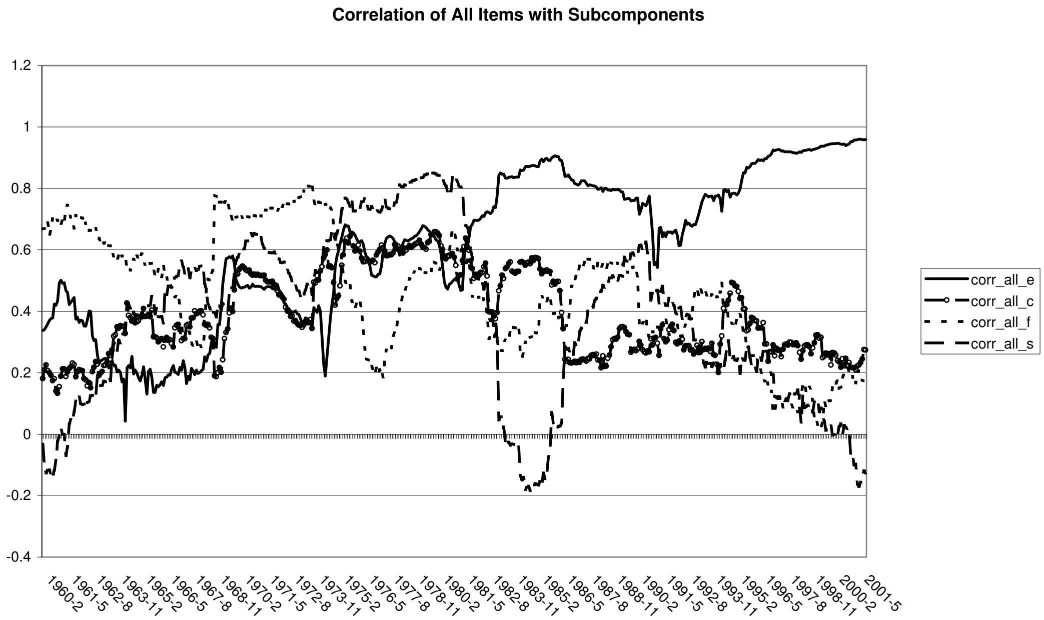

Figure 6: Correlations of US all items CPI and components

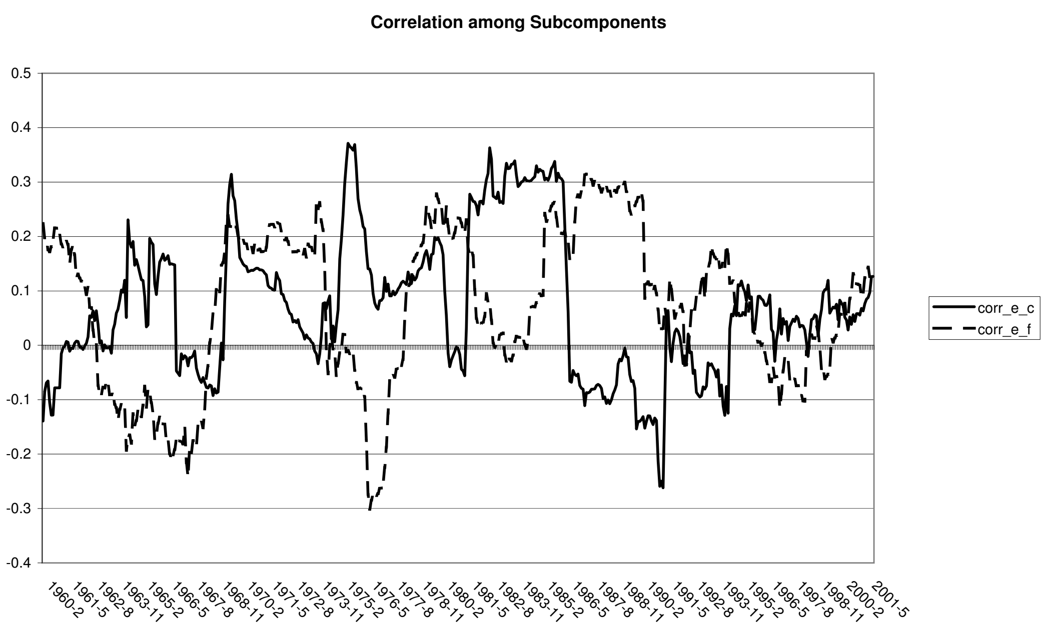

Figure 7: Correlations among US CPI components

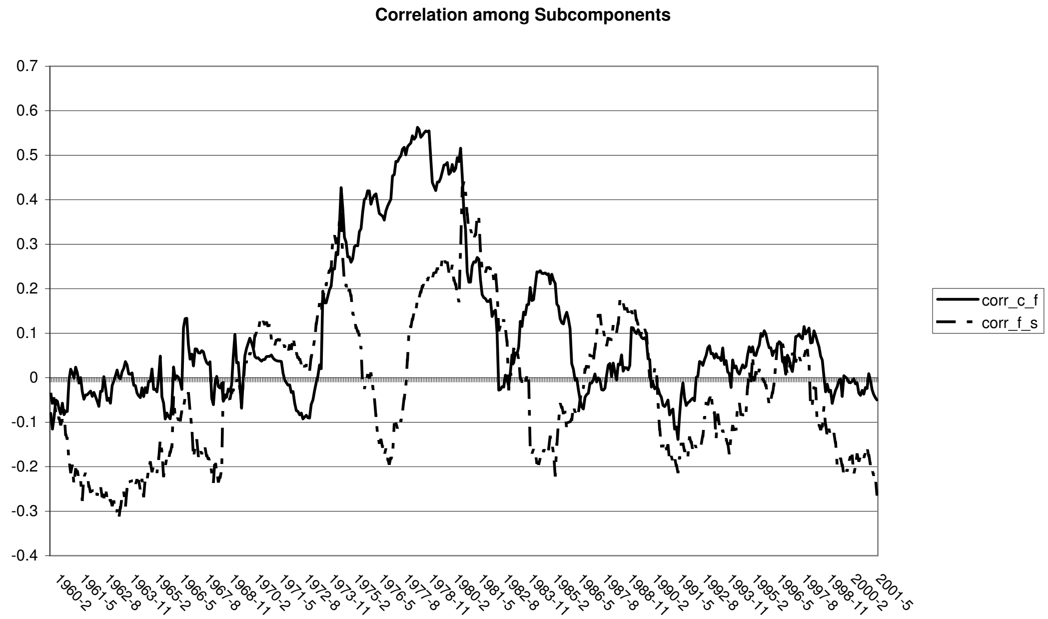

Figure 8: Correlations among US CPI components

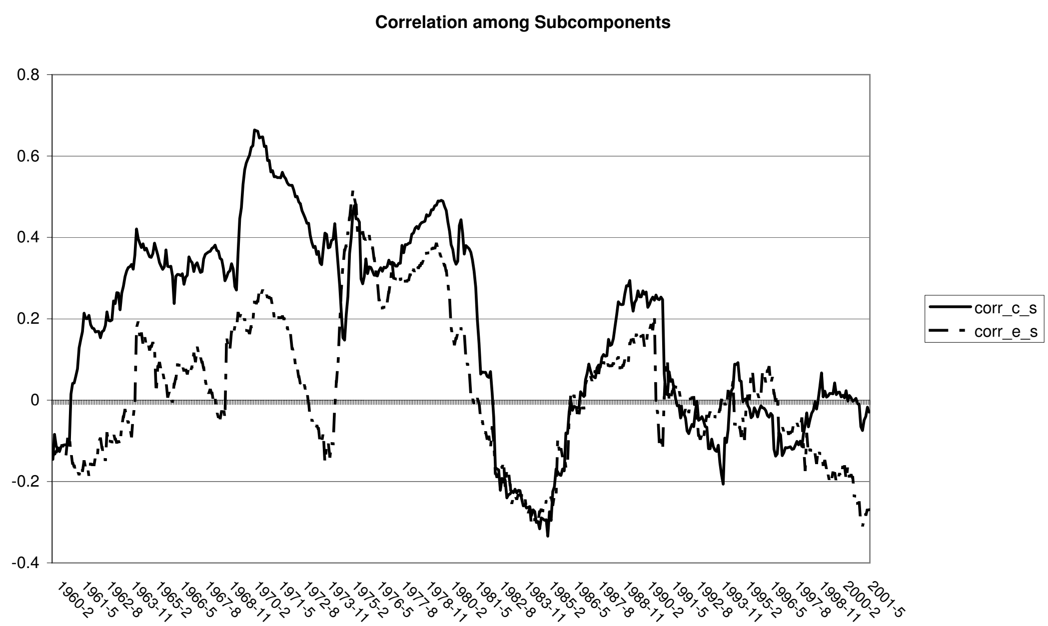

Figure 9: Correlations among US CPI components

Figure 10: Weights of US CPI components

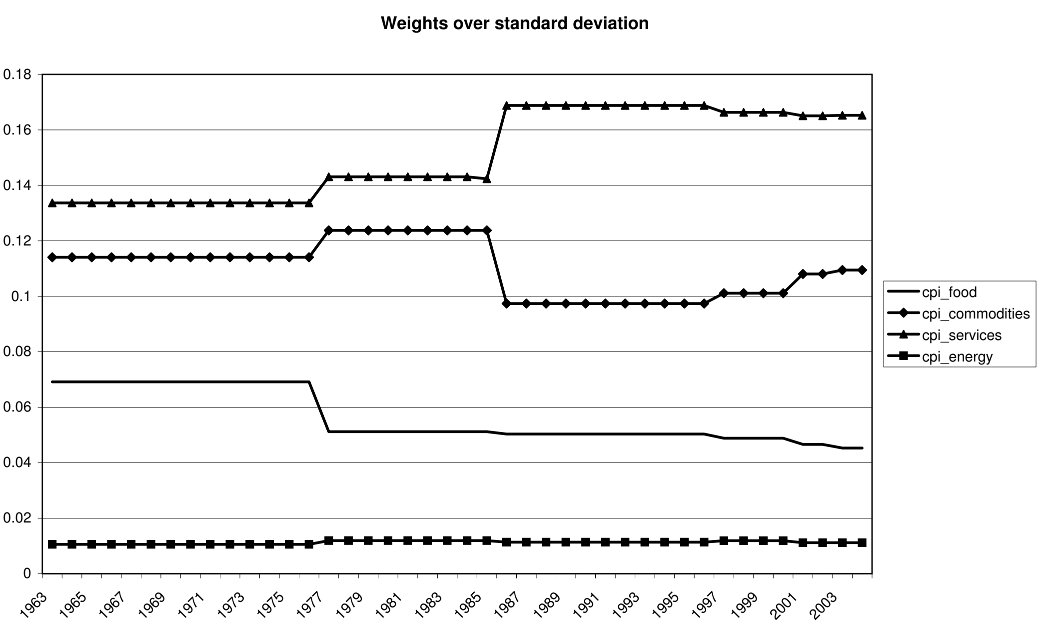

Figure 11: Weights of US CPI components relative to their standard deviation

Table 14: **Ratio Percentage Change in Weight and the Standard Deviation of the mom inflation for the respective CPI subcomponent**

component	CPI_f	CPI_c	CPI_s	CPI_e
1976-1977	-6.181	3.208	2.813	2.007
1985-1986	-0.429	-7.683	5.384	-0.563
1996-1997	-0.859	1.544	-0.505	0.561
2000-2001	-1.375	2.705	-0.261	-0.703
2002-2003	-0.843	0.503	0.045	0.002
2004-2005	-0.403	2.363	-0.449	-0.489

Note: Upper panel: standard deviation of CPI subcomponent over expanding sample (from 1960 up to year prior to the change weights)

Table 15: US, RMSFE ratios, 1970-1983 (Est. 1960-1969)

horizon	1	12
RMSFE AR^{SIC}	0.280	2.660
MSFE ratios over AR^{SIC}		
$FM(f1)$	0.969	1.000
$FM(f2)$	0.964	0.976
$FM(f3)$	0.972	0.979
$FM(f1)_1$	0.957	1.003
$FM(f2)_1$	0.936	0.964
$FM(f3)_1$	0.948	0.969
$FM(f1)_{SIC}$	0.979	1.003
$FM(f2)_{SIC}$	0.970	0.951
$FM(f3)_{SIC}$	0.962	0.956
p^f	1.007	0.986
p^c	0.982	1.014
p^s	1.005	0.995
p^e	1.004	0.997
p^{comb}	0.995	0.995

Note: RMSFE (not annualised) for AR$^{(SIC)}$ model in percentage points, Recursive estimation samples 1960(1) to 1970(1),...,1983(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, FM(f): factor models with 1,2,3 static factors, FM(f)$_1$: factor models with 1,2,3 factors with 1 lag, FM(f)$_{SIC}$: factor models with 1,2,3 factors with lags chosen by SIC, principal component estimators of static factors; p^f, p^c, p^s, p^e: single predictor models with respective subcomponent as predictor, p^{comb}: simple average of the forecasts with the four disaggregate component models

Table 16: US, RMSFE ratios, 1984 - 2004 (Est. 1960-1983)

horizon	1	12
RMSFE AR^{SIC}	0.193	1.296
MSFE ratios over AR^{SIC}		
$FM(f1)$	0.996	0.980
$FM(f2)$	1.007	1.009
$FM(f3)$	1.010	0.999
$FM(f1)_1$	1.002	0.972
$FM(f2)_1$	1.010	1.017
$FM(f3)_1$	0.992	0.975
$FM(f1)_{SIC}$	1.008	0.984
$FM(f2)_{SIC}$	0.999	0.996
$FM(f3)_{SIC}$	1.004	1.005
p^f	0.999	0.971
p^c	1.006	0.980
p^s	0.998	1.044
p^e	1.013	1.017
p^{comb}	1.000	0.995

Note: RMSFE (not annualised) for AR$^{(SIC)}$ model in percentage points, Recursive estimation samples 1960(1) to 1984(1),...,2004(12), Super and subscripts indicate model selection procedure, SIC: Schwarz criterion, FM(f): factor models with 1,2,3 static factors, FM(f)$_1$: factor models with 1,2,3 factors with 1 lag, FM(f)$_{SIC}$: factor models with 1,2,3 factors with lags chosen by SIC, principal component estimators of static factors; p^f, p^c, p^s, p^e: single predictor models with respective subcomponent as predictor, p^{comb}: simple average of the forecasts with the four disaggregate component models