

Forecasting euro area inflation: Does aggregating forecasts by component improve forecast accuracy? ¹

Kirstin Hubrich²

Research Department, European Central Bank

First version: July 2001, this version: November, 2002

Preliminary

Abstract

Monitoring and forecasting price developments in the euro area is essential in the light of the second pillar of the ECB's monetary policy strategy. This study analyses whether the forecasting accuracy of euro area inflation models can be improved by aggregating forecasts of subindices of the Harmonized Index of Consumer Prices (HICP) as opposed to forecasting the aggregate HICP directly. The analysis includes univariate and multivariate linear time series models and distinguishes between different forecast horizons, HICP components and inflation measures. Various model selection procedures are employed to select models for the aggregate and the disaggregate components. The results indicate that aggregating forecasts by component does not necessarily help for forecasting year-on-year inflation twelve months ahead.

JEL Codes: E31, E37, C53, C32

Keywords: euro area inflation, HICP subindex forecast aggregation, linear time series models

¹A first version of the paper was drafted while the author was researcher at the Dutch Central Bank. I thank Lutz Kilian and Peter Vlaar for helpful discussions. I also thank Peter van Els, Tohmas Karlsson, Helmut Luetkepohl as well as participants of the European Econometric Society Meeting 2002 in Venice, a seminar of the Working Group of Econometric Modelling of the ESCB and a seminar at the Dutch Central Bank for useful comments. The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank.

²European Central Bank, Research Department, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; e-mail: kirstin.hubrich@ecb.int

1 Introduction

The primary objective of the ECB's monetary policy is price stability. Price stability has been defined by the Governing Council of the ECB as "a year-on-year increase in the Harmonized Index of Consumer Prices (HICP) for the euro area of below 2%" (European Central Bank, 2001, p.38).

The European System of Central Banks (ESCB) is monitoring and projecting prices under the second pillar of the ECB's monetary policy strategy to assess price developments in the euro area. Since December 2000 the ECB has been publishing its inflation projection for the euro area.³ Further insights regarding the performance of different forecasting strategies for euro area inflation are highly relevant for policy makers and ECB observers.

In the context of forecasting euro area inflation the question arises to what extent the forecasting accuracy of different time series models for aggregate inflation can be improved by modelling inflation of disaggregated subcomponents and aggregating forecasts based on these models. Contemporaneous (dis-)aggregation can be considered in two dimensions: the aggregation of national forecasts for euro area countries and the aggregation of HICP subcomponent forecasts. The forecasting accuracy of aggregating country-specific forecasts in comparison with forecasts based on aggregated euro area wide data has been analysed on the basis of a broad range of models in Marcellino, Stock & Watson (2002). Other studies have focused on specific methods incorporating national information into forecasts of euro area wide inflation. For example, Angelini, Henry & Mestre (2001) and Cristadoro, Forni, Reichlin & Veronese (2001) employ dynamic factor models in this context.

In contrast to these studies, the aim of this analysis is to compare the forecasting accuracy of models forecasting the aggregated HICP directly as opposed to aggregating forecasts for HICP subcomponents. A broad range of models based on various model selection procedures is employed. The comparison is based on data for the euro area as a whole relevant for the monetary policy of the ECB.⁴

³The word 'projection' in contrast to forecasting is used by the ECB to indicate that the published projections are based on a set of underlying technical assumptions, including the assumption of unchanged short-term interest rate. In contrast, no assumptions for the development of any of the variables over the forecast horizon are made in this study.

⁴Espasa, Senra & Albacete (2002) compare the forecasting accuracy of aggregate and disaggregate approaches of forecasting euro area inflation, disaggregating either by coun-

The debate about aggregation versus disaggregation of economic variables goes back to Theil (1954) and Grunfeld & Griliches (1960). There are two main arguments for aggregating forecasts for disaggregated variables instead of forecasting the aggregate variable of interest directly. One rationale is that the disaggregated variables can be better modelled and, therefore, predicted more accurately than the aggregate variable for two reasons: First, modelling disaggregated variables would involve using a larger information set and, second, specifications may vary across the different disaggregated variables (see Barker & Pesaran, 1990b). Aggregating forecasts of the disaggregate components would in turn imply more accurate prediction of the aggregate in terms of mean square forecast error. A second intuitive argument in favour of disaggregation is that forecast errors of disaggregated components might cancel partly, leading to higher aggregated prediction accuracy (see also Clements & Hendry (2002b) for a discussion on forecast combination as a bias correction). However, it can also be argued that it is actually better to forecast the aggregate directly. Since the models for the disaggregate variables will in practice not be perfectly specified (see Grunfeld & Griliches, 1960), the misspecified disaggregate model might not improve the forecast accuracy for the aggregate, especially in the presence of unexpected shocks to some of the disaggregate variables as will be seen in the analysis presented in this study. Furthermore, unexpected shocks might affect the forecast errors of some of the disaggregate variables in the same direction so that forecast errors do not cancel.

In this study, I examine whether aggregating inflation forecasts based on HICP subindices is really better than forecasting aggregate HICP inflation directly. I compare both approaches to forecasting euro area inflation under different aspects that have emerged from earlier literature based on asymptotic theory and from Monte Carlo simulations to determine the effect of disaggregation on forecast accuracy. The main aspects include i) different forecast models, ii) different model selection procedures, iii) different forecast horizons, and iv) different inflation measures (e.g. aggregate 'headline' inflation including all subindices and HICP inflation excluding unprocessed food and energy prices, sometimes referred to as 'core' inflation). The forecasting methods include a random walk model for year-on-year inflation,

tries or by sectors. Differences between their results and the results presented in the current study arise mainly from different models employed and a different transformation of HICP underlying the evaluation of the relative forecast accuracy.

univariate autoregressive models and vector autoregressive models based on various model selection strategies. Univariate and multivariate linear time series models are chosen for the comparison since these are often used for forecasting inflation in Europe on a national or euro area wide level. Vector error correction models have not been included in the comparison since those models can fail badly in forecasting in the presence of structural breaks in the equilibrium mean (see e.g. Clements & Hendry (2002a)).⁵ Nonlinear time series models are not considered here due to the short time series available for estimation.⁶ Time varying parameter models are not considered here, either. Stock & Watson (1996) suggest that gains from using time varying parameter models for forecasting are generally small or non-existent, especially on short horizons.⁷ Various model selection strategies are employed in this study to select models for the aggregate HICP and its disaggregate components. These include choosing an information set guided by economic theory where the same model specification is chosen for each of the subcomponents. Furthermore, the Schwarz criterium selecting parsimonious univariate models and a general-to-specific modelling strategy implemented in the software package PcGets (Hendry & Krolzig, 2001a) are employed.⁸ The latter model selection procedures allow for varying specification across subcomponents in terms of lag order and / or variables included.

The remainder of the paper is structured as follows: In the second section I discuss some asymptotic and small sample simulation results from the literature regarding the relative forecasting performance of aggregated forecasts of time series subcomponents. The third section presents the data, forecast methods and the model selection procedures on which the forecast comparison is based. Furthermore, the results for the relative forecast accuracy of the aggregated versus the disaggregated approach to forecasting euro area inflation are discussed. Finally, section four draws some tentative conclusions from the analysis and points out directions for further research.

⁵For an application of a VECM to forecasting euro area inflation taking into account a cointegration relation between HICP subindices, see Espasa et al. (2002).

⁶An exposition of the forecasting performance of non-linear models can, for example, be found in Clements & Hendry (1999, Ch.10).

⁷See Canova (2002) for a recent more favourable evaluation of the forecasting performance of Bayesian time varying parameter models.

⁸For an analysis of the role of model selection strategies on forecasting failure, see e.g. Clements & Hendry (2002a).

2 Forecasting contemporaneously aggregated time series: Some results from the literature

In empirical analysis the researcher often has to work with temporally or contemporaneously aggregated variables. Recently, there has been renewed interest in the consequences of temporal aggregation for empirical analysis (see Marcellino (1999)). Similarly, the effects of contemporaneous aggregation across national variables in the context of modelling euro area developments have received increasing interest.⁹ The focus of this study is on analysing the effects of contemporaneous aggregation of subcomponents of time series variables on forecasting accuracy which in the empirical literature has found rather limited attention so far.

Consider forecasting a contemporaneously aggregated variable that is defined as a variable consisting of the sum or the weighted sum of a number of different disaggregated subcomponents at time t . The contemporaneous aggregate can be written as

$$y_t^{agg} = w_1 y_t^1 + w_2 y_t^2 + \dots + w_n y_t^n, \quad t = 1, \dots, T,$$

where y_t^j ($j = 1, \dots, n$) are the subcomponents of y_t^{agg} , n is the number of subcomponents considered and w_j , $j = 1, \dots, n$, are the aggregation weights. It is assumed that the aggregation weights are fixed, i.e. they do not change over time.¹⁰ Thus, y_t^{agg} is assumed to be a linear transformation of the stochastic processes y_t^j . Two different forecasts of the aggregate will be considered in this study. The direct forecast of the aggregated variable, denoted as \hat{y}_t^{agg} , and an indirect forecast of the aggregated variables by aggregating the n subcomponents forecasts \hat{y}_t^j ($j = 1, \dots, n$), i.e. $\hat{y}_{subF,t}^{agg} = \sum w_j \hat{y}_t^j$.

The issue of contemporaneous aggregation of economic variables has already been discussed and analysed in an early contribution by Theil (1954) who argues that disaggregation pays for model specification. Grunfeld & Griliches (1960), however, point out that if the micro equations are not as-

⁹In addition to the papers on forecasting inflation in the euro area mentioned in the introduction, see e.g. Zellner & Tobias (2000) on disaggregation and forecasting performance in industrialized countries.

¹⁰For an extension allowing for time varying weights, see e.g. Lewbel (1992) and Van Garderen, Lee & Pesaran (2000).

sumed to be perfectly specified, aggregation is not necessarily bad since the specification error might be higher than the aggregation error.

In the course of further developments in this discussion, the theoretical econometric literature on contemporaneous aggregation of time series has focussed on several themes. One strand of the theoretical literature has concentrated on deriving the nature of the data generating process (DGP) of the aggregated process if the subcomponents are assumed to follow a certain DGP (e.g. Rose (1977) for ARIMA processes, Lippi & Forni (1990) for ARMAX models and Nijman & Sentana (1996) for GARCH models).

Another strand of the theoretical literature has focussed on the effect of contemporaneous aggregation on forecasting accuracy, for example Rose (1977), Tiao & Guttman (1980), Wei & Abraham (1981), Kohn (1982) and Lütkepohl (1984a, b, 1987). Leamer (1990) derives an optimal degree of disaggregation in terms of the prediction error.¹¹

In the following, the main asymptotic and small sample simulation results from the latter strand of the literature, that are of interest in the context of the empirical analysis presented later, are summarized. Based on asymptotic theory the following results can be derived. If the DGP is known, aggregating subcomponent forecasts is better in terms of a mean square forecast error (MSFE) criterion than forecasting the aggregate directly, $MSFE(\hat{y}_{subF}^{agg}) < MSFE(\hat{y}^{agg})$. However, the usefulness of this result is limited because in practice the DGP is usually not known. If the assumption of a known data generation process (DGP) is relaxed and it is assumed that the process order is finite and estimated using a consistent order selection criterion, the aggregation of forecasts of the components can actually be inferior to forecasting the aggregated time series directly, $MSFE(\hat{y}_{subF}^{agg}) > MSFE(\hat{y}^{agg})$. Therefore, asymptotic theory provides inconclusive results regarding the ranking of the disaggregate and the aggregate approach to forecasting the variable of interest.

Despite the effort to understand the theoretical aspects of the effect of disaggregation on forecasting, this line of research has yielded few practically useful insights. Therefore, Lütkepohl (1984a, 1987) presents Monte Carlo simulations to analyse the relative small sample accuracy in terms of the MSFE of directly forecasting the aggregate and aggregating subcompo-

¹¹Some related issues and results are presented in a paper by Granger & Morris (1976). Granger (1990) provides a survey on aggregation of time series variables. Further papers, including a number of empirical studies, can be found in Barker & Pesaran (1990a).

nent forecasts. The small sample simulations largely confirm the asymptotic results. He finds that the small sample ranking of the two approaches are mixed. The results suggest that it is not necessarily better to aggregate the subcomponent forecasts instead of forecasting the aggregate. If the subcomponents are uncorrelated and the forecast horizon is short, then aggregating the subcomponent forecasts may lead to a lower MSFE for certain DGPs.

Overall, results from asymptotic theory and small sample simulations do not seem to give a clear answer as regards the relative forecast accuracy of the disaggregate versus the aggregate forecasting approach. Therefore, it seems that whether aggregation of subcomponent forecasts improves forecast accuracy is largely an empirical question. Furthermore, the theoretical analyses as well as the small sample simulation assume certain DGPs. In practice, however, the DGP is not known. Therefore, an empirical out-of-sample experiment is carried out in this study to gather insights about the effect of contemporaneous aggregation on forecasting accuracy for the actual DGP of euro area HICP that is at the center of interest of monetary policy in the euro area.

Previous simulation studies suggest that the performance in small samples of the two approaches to forecast aggregated time series, i.e. forecasting the aggregated series directly or aggregating the forecasts of its subcomponents, will depend on many different factors: the forecasting method, the dimension of the model employed, i.e. the use of univariate or multivariate models and the number of variables included, the model selection procedure employed, the forecast horizon, the degree of dependence between the component series, and, finally, the sample size. Those factors will be investigated in the context of forecasting euro area inflation.

The theoretical literature and the simulation studies on the effects of contemporaneous aggregation on the forecasting accuracy mainly focus on univariate models. Multivariate models are only considered with subcomponents as endogenous variables, but no explanatory variables are introduced (e.g. Lütkepohl, 1987). In the out-of-sample forecasting experiment presented in the current study, this set-up will be extended including a larger set of economic variables.

Most of the asymptotic results either assume the true DGP to be known or correctly specified, or they are derived assuming that a consistent model selection criterion is used. In practice, estimation variability will be a main factor in reducing the relative forecasting efficiency of models with a high number of parameters. Therefore, the model selection procedure is impor-

tant in deriving the final model. Lütkepohl (1984a) presents results for subset VARs that lead to lower estimation variability due to parsimonious specification. Another possibility to choose a parsimoniously specified model is to use the general-to-specific model selection procedure suggested by Hendry & Krolzig (2001b) and implemented in PcGets (Hendry & Krolzig, 2001a). This model selection procedure has been included in the comparison presented below.

3 Simulated out-of-sample forecast comparison

To evaluate the relative forecast accuracy of forecasting aggregate HICP directly versus aggregating the forecasts of HICP subcomponents, a simulated out-of-sample forecast experiment is carried out. One to twelve step ahead forecasts are performed based on different linear time series models estimated on recursive samples. The main criterion for the comparison of the forecasts employed in this study, as in a large part of the literature on forecasting, is the root mean square forecast error (RMSFE).

The data include aggregated overall HICP for the euro area as well as its breakdown into five subcomponents: unprocessed food, processed food, industrial goods, energy and services prices. A range of explanatory variables for inflation is also considered.

The data used in the analysis for the variables possibly included in the models are presented in section 3.1. The forecast methods and model selection procedures employed are outlined in section 3.2, whereas section 3.3 presents the results.

3.1 Data

The data employed are of monthly frequency¹², starting in 1992(1) until 2001(12). This is a relatively short sample, which is determined by the availability of data. Model selection and estimation is carried out on the basis of 12 recursive samples starting from 1992(1) up to 2000(1), extending the sam-

¹²Except for unit labour costs which are of quarterly frequency and have been interpolated.

ple by one month sequentially. Seasonally adjusted data have been chosen¹³ because of the changing seasonal pattern in some of the HICP subcomponents for some countries due to a measurement change.¹⁴ The sensitivity of the results toward using seasonally unadjusted data has been analysed on the basis of a shorter sample. The results show no substantial change in the conclusions. The breakdown of the HICP (sub-)indices is chosen in accordance with the data published in the ECB Monthly Bulletin. The notation will be the following: HICP unprocessed food will be denoted p^{uf} , HICP processed food p^{pf} , HICP industrial production p^i , HICP energy p^e and HICP services p^s . Furthermore, aggregate HICP will be denoted p^{agg} .

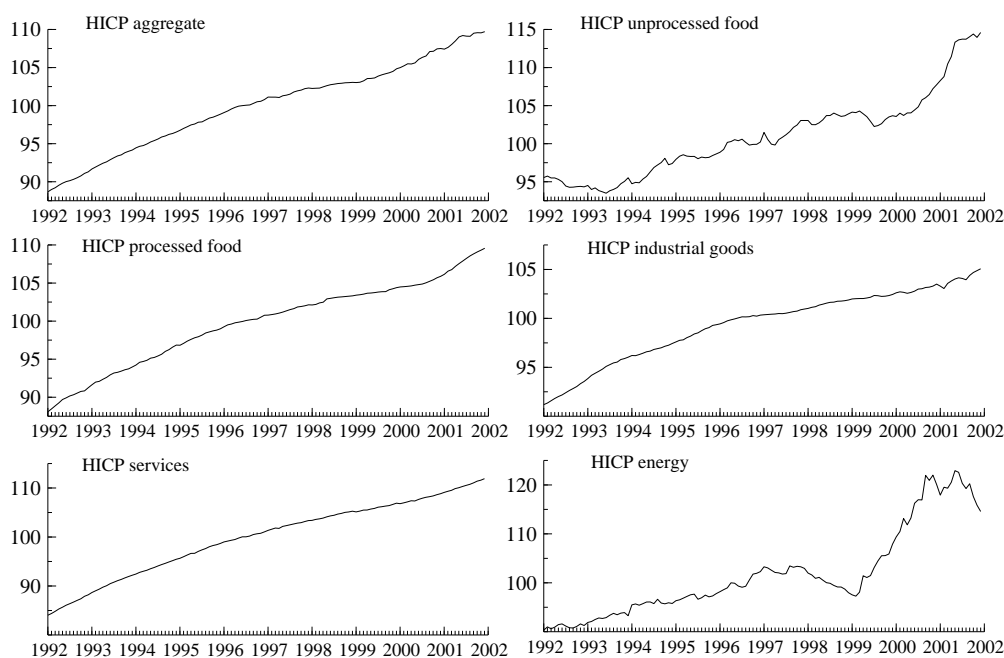


Figure 1: HICP (sub-)indices (in logarithm)

The aggregate HICP price index and the HICP subindices in logarithm are presented in Figure 1 and the year-on-year inflation rates in % of the

¹³Except for interest rates, producer prices and HICP energy that do not exhibit a seasonal pattern.

¹⁴The data are taken from the ECB and Eurostat. For details, see the Appendix on the data.

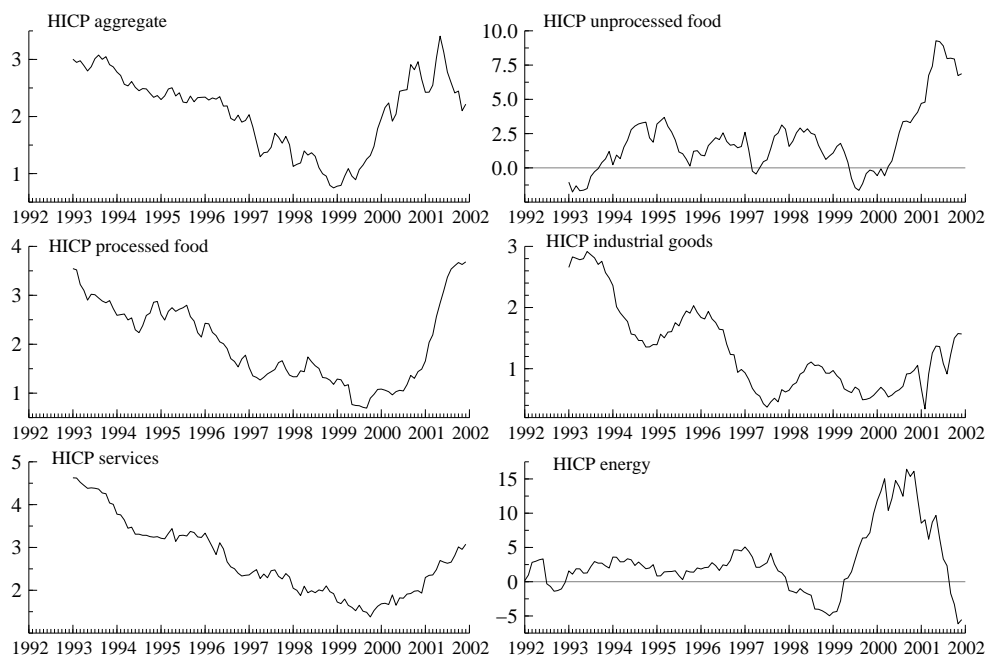


Figure 2: Year-on-year HICP inflation (in %), aggregate and subindices

respective indices are depicted in Figure 2. The HICP indices in first differences are displayed in Figure 9 in the Appendix. Aggregate HICP, HICP processed food, HICP industrial production and HICP services display a relatively smooth upward trend. In contrast, HICP unprocessed food and HICP energy exhibit a much more erratic development (see Figure 1). The annual inflation rates (see Figure 2) exhibit a downward trend for total HICP, processed food prices, prices of industrial goods and service prices roughly until 1999. Unprocessed food and energy prices do not show a downward trend, but a sharp increase in 1999 due to oil price increases and animal diseases. Augmented Dickey Fuller (ADF) tests have been carried out for all HICP (sub-)indices (in logarithm) based on the sample 1992(1) to 2000(12). The tests indicate that all (sub-)indices are integrated of order one, except the aggregate HICP and HICP services. The first differences of those two series, however, are found to be trend stationary for a sample period up to 2000(8). Therefore and because of the low power of the ADF test both aggregate HICP and HICP services are assumed to be I(1) in the analysis. A robustness analysis could investigate whether the results change if aggregate and services prices are assumed to be I(2) over all recursive samples, i.e. modelled in second differences.

Further variables that enter the large VAR model included in the forecast accuracy comparison are industrial production, y , and nominal money M3, m , producer prices, p^{prod} , import prices, p^{im} , unemployment, u , unit labour costs, ucl , commodity prices (excluding energy) in euro, p^{com} , oil prices in euro, p^{oil} , the nominal effective exchange rate of the euro, $NEER$, as well as a short-term and a long-term nominal interest rate, i^s and i^l . This choice of variables to enter the multivariate model tries to strike a balance between including relatively few variables due to the short data series available for the euro area on the one hand and including the key variables that influence inflation according to economic theory. All variables except the interest rates are in logarithms. The graphs displaying these explanatory variables are presented in the Appendix.

3.2 Forecast Methods and Model Selection

Five different forecasting models using different forecast models and different model selection procedures are employed. The random walk with drift (RW) is employed as a benchmark model. A univariate autoregressive (AR) model is included in the comparison where the lag order is parsimoniously

chosen using the Schwarz criterion, denoted AR^{SC} . Furthermore, a simple Phillips curve model, as e.g. in Stock & Watson (1999), is employed including inflation and the change in unemployment in the VAR with 2 and, alternatively, with 12 lags. Those models will be denoted $VAR^{Ph(2)}$ and $VAR^{Ph(12)}$. The fourth model is a larger VAR with a number of domestic and international variables described in the data section above, allowing for 2 lags ($VAR^{Int(2)}$). And, finally, a general-to-specific model selection strategy is employed to choose a VAR ($VAR_{Gets}^{Int(2)}$) implemented in the computer package PcGets by Hendry & Krolzig (2001a) where the choice of variables and lag length is based on mis-specification tests, structural break tests, t- and F-block tests, encompassing tests and information criteria.¹⁵ The model is selected starting with a VAR including a large potential number of domestic and international variables. For all other models (except the $VAR_{Gets}^{Int(2)}$) the model choice and simulated out-of-sample experiment are carried out using GAUSS. All models are newly estimated and the model selection procedures are carried out for each of the recursive samples.

3.3 Relative forecast accuracy

The forecasts produced by the respective method have to be transformed, since the forecast accuracy is to be evaluated in terms of root mean square forecast error (RMSFE) of year-on-year inflation. Note that the multi-horizon MSFEs do not allow comparison between different representations of the same system (Granger & Newbold, 1986). Therefore, it is important to note that here the focus is on the comparison of all HICP (sub-)indices in terms of their forecast accuracy for year-on-year inflation rates since those are most relevant from a monetary policy perspective.

The aggregated HICP is a weighted chain index, where the weights change each year. Since the end of all recursive estimation samples is in 2000, the aggregation of the forecasts is carried out using the HICP subcomponent weights of the year 2000 (at prices of 1999(12)) which would be known to the forecaster in real time. The forecasts from the models in first differences are recalculated to level forecasts and rebased to the month 1999(12) in accordance with the weights used. The weighted sum of the subcomponents forecasts is then rebased to the base year 1996 of the actual aggregate index

¹⁵A 'liberal' selection strategy has been chosen implying a higher probability of retaining relevant variables at the risk of retaining irrelevant ones.

and transformed into year-on-year inflation rates. Those are then compared with the respective realization of year-on-year inflation. The actual weights of 2000 used are 8.2 % for unprocessed food, 12.6 % for processed food, 32.6 % for industrial goods, 9.0 % for energy and 37.6 % for services prices.¹⁶

Table 1 presents the comparison of the relative forecast accuracy measured in terms of RMSFE of year-on-year inflation of the direct forecast of aggregate inflation ($\Delta_{12}\hat{p}^{agg}$) and the aggregated forecasts of the subindices ($\Delta_{12}\hat{p}_{subF}^{agg}$).

Table 1: **Relative forecast accuracy, RMSFE of year-on-year inflation in percentage points, Recursive estimation samples 1992(1) to 2000(1),...,2000(12)**

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$
<i>RW</i>	0.181	0.182	0.366	0.347	0.492	0.460
<i>AR^{SC}</i>	0.213	0.208	0.433	0.608	0.590	0.949
<i>VAR^{Ph(2)}</i>	0.201	0.200	0.604	0.603	0.950	0.955
<i>VAR^{Ph(12)}</i>	0.211	0.221	0.390	0.395	0.864	0.789
<i>VAR^{Int(2)}</i>	0.156	0.142	0.520	0.571	0.735	0.800
<i>VAR^{Int(2)}_{Gets}</i>	0.167	0.165	0.590	0.623	0.925	1.0

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph: Phillips curve model including inflation and unemployment, Int: model including international variables in addition to domestic ones, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest RMSFE per column

Since different forecast horizons might lead to different rankings of the forecasting methods, the comparison is carried out for short-term to medium-term forecast horizons, 1 to 12 months ahead. The RMSFE evaluation is based on recursive forecasts that involve an average of the respective horizon forecasts over all 12 recursive samples. In the paper, the results for 1-,6- and 12-months ahead forecasts are presented. The one step ahead forecasts are starting with the forecast for 2000(2) based on the estimation sample

¹⁶Changes in weights from year to year are relatively small.

1992(1) to 2000(1), the second forecast is for 2000(3) based on the estimation sample up to 2000(2), etc., the 12th forecast for 2001(1) is then based on the estimation sample up to 2000(12). Similarly, 12-period-ahead forecasts are carried out for 12 different estimation samples. The forecast for 2001(1) is based on the sample up to 2000(1), whereas the last 12 step ahead forecast is carried out for 2001(12) based on the estimation sample until 2000(12).

Another simulated out-of-sample experiment has been carried out extending the forecast evaluation period backwards to analyse the sensitivity of the results towards a specific forecast period. Recursive estimation then started with a sample from 1992(1) to 1998(1). The results of this analysis did not change the conclusion of the paper.¹⁷ The following presentation focusses on the shorter sample period.

For a 1-step-ahead forecast horizon aggregating subcomponent forecasts tends to outperform forecasting the aggregate directly in terms of RMSFE (see Table 1). Whereas for the *RW* and the $VAR^{Ph(2)}$ both approaches show almost the same performance, for most of the other models aggregating the subcomponent forecasts performs better. The large $VAR^{int(2)}$ and the $VAR^{int(2)}_{Gets}$ perform best overall. These models are probably better in capturing the increase in energy prices and its second round effects on the other price components as well as the increase unprocessed food prices in 2000 by explicitly including oil prices, commodity prices and producer prices, among others.

In contrast, for a forecast horizon of 6 and 12 months, directly forecasting aggregate inflation tends to perform better in RMSFE terms. The *RW* turns out to be best overall for the period considered and the AR^{SC} model is second. It should be noted that in contrast to most other models the *RW* exhibits a better performance for aggregating subcomponent forecasts.

However, the MFE in Table 2 shows that the modulus of the bias of the forecast is higher for those models that show a lower RMSFE in the case of aggregating the subcomponent forecasts ($\Delta_{12}\hat{p}_{subF}^{agg}$) for a one month horizon. For a 12 months forecast horizon the modulus of the bias for the AR^{SC} of the direct forecast and for $VAR^{Ph(12)}$ of the indirect forecast is lower than the for the *RW* that showed the lowest RMSFE.

It should be noted that the general-to-specific model selection procedure for $VAR^{int(2)}_{Gets}$ does improve in RMSFE terms over the simple Phillips curve model, but not over the large $VAR^{int(2)}$ for 1 month ahead forecasts. For

¹⁷The results are available from the author upon request.

Table 2: **Relative forecast accuracy, MFE of year-on-year inflation in percentage points, Recursive estimation samples 1992(1) to 2000(1),...,2000(12)**

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}_{subF}^{agg}$
<i>RW</i>	0.013	-0.007	0.260	-0.231	0.314	-0.251
<i>AR^{SC}</i>	-0.010	-0.048	0.213	-0.467	0.282	-0.734
<i>VAR^{Ph(2)}</i>	0.033	-0.037	0.492	-0.489	0.792	-0.772
<i>VAR^{Ph(12)}</i>	-0.013	0.012	-0.005	-0.129	-0.312	-0.0005
<i>VAR^{Int(2)}</i>	0.017	-0.031	0.303	-0.339	0.439	-0.487
<i>VAR^{Int(2)}_{Gets}</i>	-0.039	-0.054	-0.440	-0.528	-0.713	-0.872

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph: Phillips curve model including inflation and unemployment, Int: model including international variables in addition to domestic ones, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicates lowest absolute MFE per column

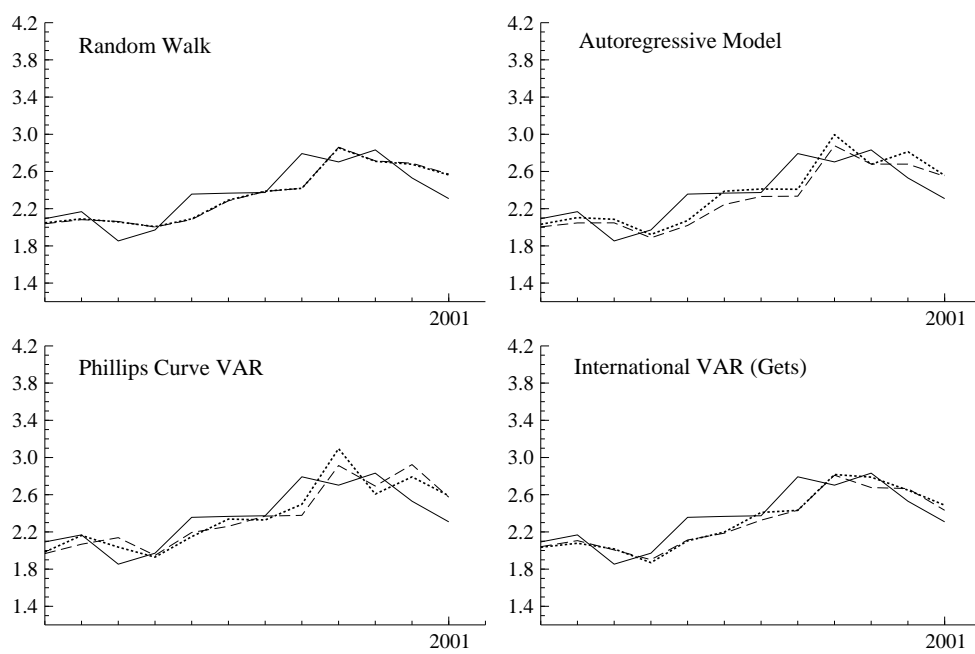


Figure 3: Year-on-year inflation rate and forecasts in %, 1 month ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

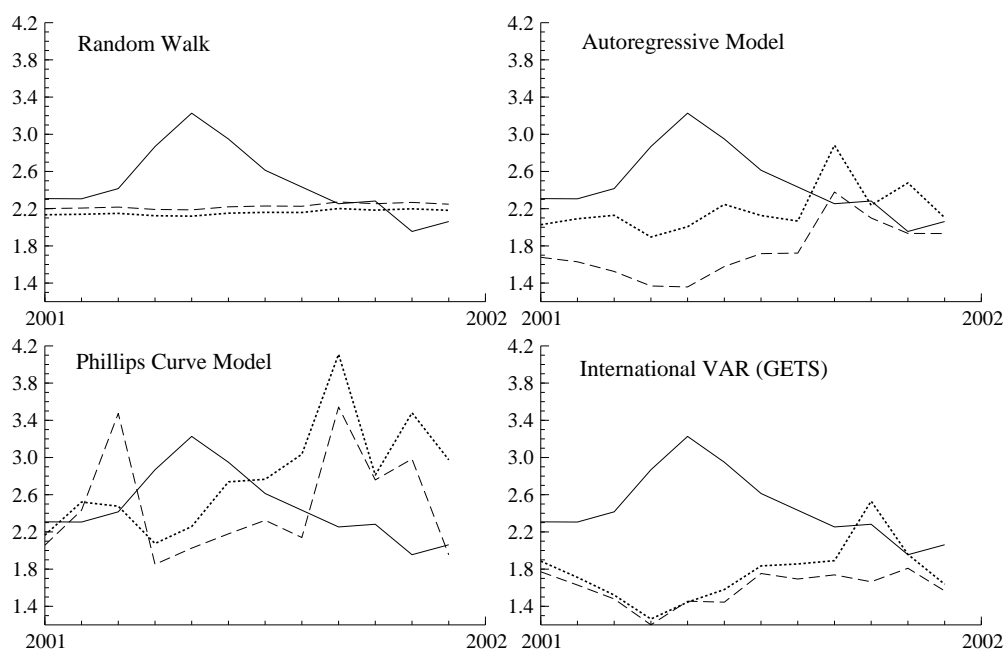


Figure 4: Year-on-year inflation rate and forecasts in %, 12 months ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

a twelve months ahead forecast, the $VAR_{Gets}^{Int(2)}$ does not gain anything in RMSFE terms over the $VAR^{Ph(12)}$ in forecasting the aggregate directly, but loses in aggregating the subcomponent forecasts. This indicates that the model selection procedure that improves the in-sample fit does not necessarily improve out-of-sample forecast accuracy. This has also been pointed out by e.g. Clements & Hendry (2002a).

To evaluate how good or bad these models are in terms of predicting year-on-year inflation and how much the direct forecast of the aggregate actually differs from the aggregated subcomponent forecast based on the same method, both forecasts are presented graphically for each model together with the respective realization.

Figure 3 presents the actual year-on-year inflation rates and the 1-step ahead forecasts for 12 recursive samples for the RW , the AR , the $VAR^{Ph(12)}$ and the VAR_{Gets}^{int} . It shows that for a one step ahead forecast horizon there is hardly any difference between the direct and indirect approach to forecasting year-on-year inflation. The largest difference is for $VAR^{Ph(12)}$ of about 0.2 percentage points.

For a forecast horizon of 12 months, which is more relevant for monetary policy, a similar result can be seen in Figure 4 for the RW . For those models there is hardly any difference between the direct and indirect forecast whereas the difference for AR^{SC} , the $VAR^{Ph(12)}$ and the $VAR_{Gets}^{Int(2)}$ is between 0.1 up to almost 2 percentage points. For the AR^{SC} and the $VAR_{Gets}^{Int(2)}$ the RMSFE indicates a better performance of forecasting the aggregate year-on-year inflation directly. It is worth noting that these are the models where the specification varies across price subcomponents. The predictive failure of all models for the 12 months ahead forecast over most of the recursive samples can be explained by their failure to predict several unexpected shocks: The increase in year-on-year changes of unprocessed food prices in the first half of 2001 due to the effects of animal diseases (BSE and Foot-and-Mouth disease); the increase in year-on-year changes of processed food prices over the whole year 2001 due to lagged effects of the animal diseases coming from unprocessed food prices; the increase in year-on-year changes of industrial goods prices is to a large extent due to lagged effects of the increase of energy prices and the depreciation of the euro. Furthermore, the decline in the year-on-year change of energy prices in the second half of 2001 is not well captured by either of the models.

Figures 5 and 6 show the results for the RW and the AR model for each of

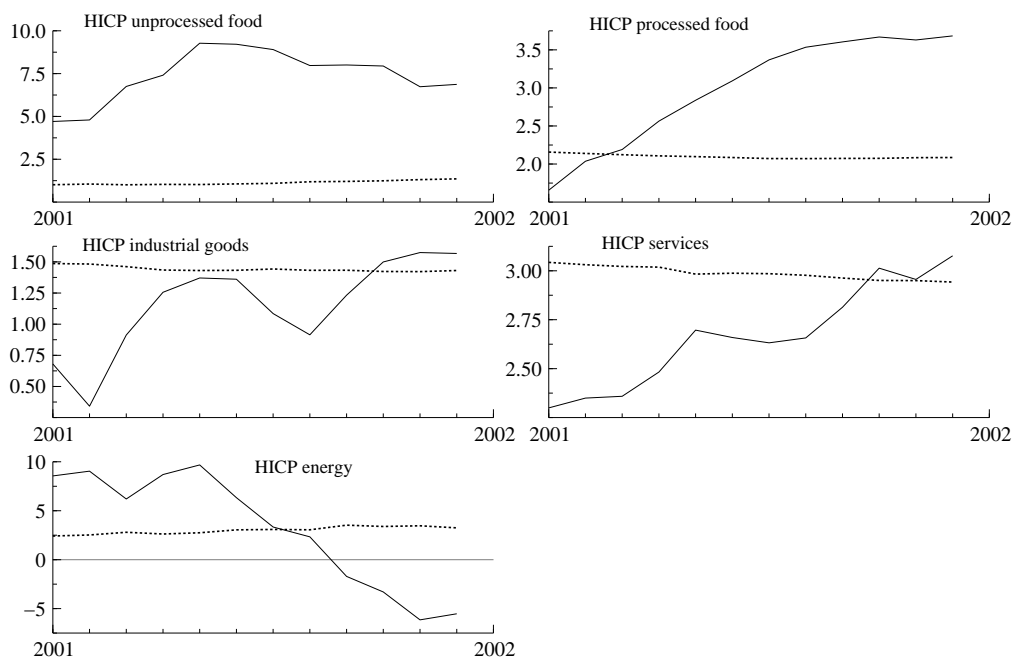


Figure 5: Year-on-year inflation rate in %, solid: actual, dashed: RW sub-component forecast, 12 months ahead

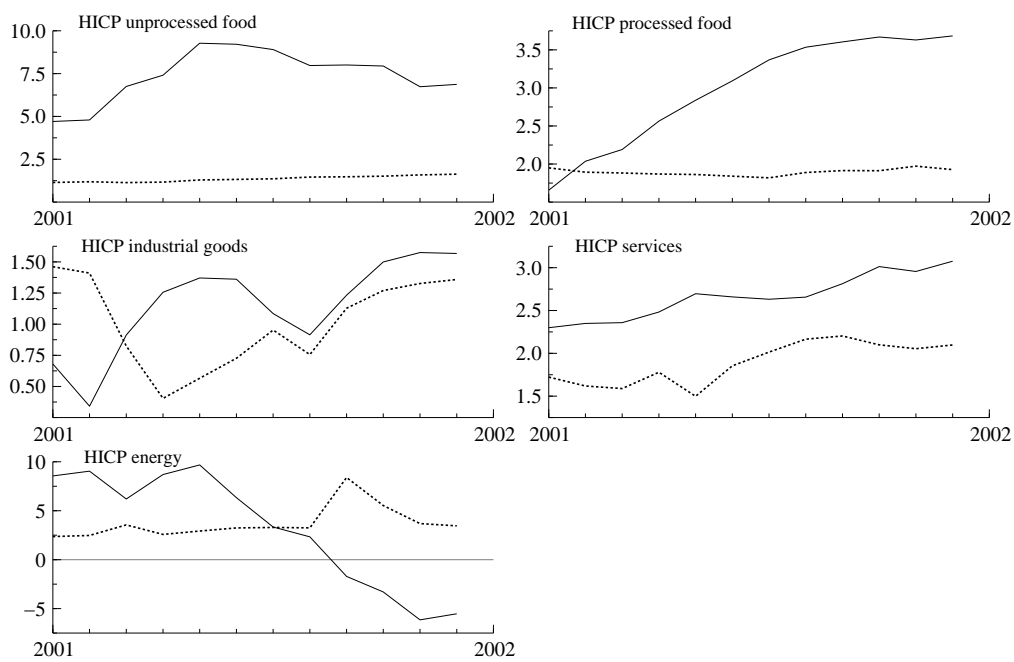


Figure 6: Year-on-year inflation rate in %, solid: actual, dashed: AR sub-component forecast, 12 months ahead

the subcomponents. Figure 6 shows that unprocessed food, processed food and services prices are substantially under-predicted for the whole year of 2001 and energy prices for the first half of the year 2001. A similar picture arises for all other models (except the *RW*, see Figure 5, where industrial goods prices and services prices are over-predicted for most of 2001). This might provide an explanation why aggregating subcomponent forecasts is not better than forecasting the aggregate inflation rate directly: The subcomponents are affected by unexpected shocks in the same way and therefore lead to forecast failures in the same direction.

Overall, there is a tendency for a better forecast accuracy of forecasting aggregate year-on-year inflation directly over longer horizons, especially for a 12 months horizon of interest for monetary policy.

Table 3: Relative forecast accuracy, RMSFE of year-on-year inflation of HICP excluding unprocessed food and energy in percentage points, Recursive estimation samples 1992(1) to 2000(1),...,2000(12)

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}_{subF}^{core}$	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}_{subF}^{core}$	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}_{subF}^{core}$
<i>RW</i>	0.080	0.084	0.273	0.297	0.387	0.422
<i>AR^{SC}</i>	0.064	0.051	0.247	0.205	0.635	0.544
<i>VAR^{Ph(2)}</i>	0.065	0.060	0.273	0.269	0.677	0.674
<i>VAR^{Ph(12)}</i>	0.056	0.047	0.247	0.262	0.716	0.737
<i>VAR^{Int(2)}</i>	0.095	0.091	0.285	0.266	0.596	0.491
<i>VAR^{Int(2)}_{Gets}</i>	0.065	0.074	0.292	0.248	0.731	0.572

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph: Phillips curve model including inflation and unemployment, Int: model including international variables in addition to domestic ones, Gets: model selection with PcGets (Hendry & Krolzig, 2001a)

Another aggregate inflation measure that is of interest for the ECB is HICP inflation excluding energy and unprocessed food, sometimes referred to as 'core' inflation. The results in terms of the RMSFE of year-on-year 'core' inflation are presented in Table 3.

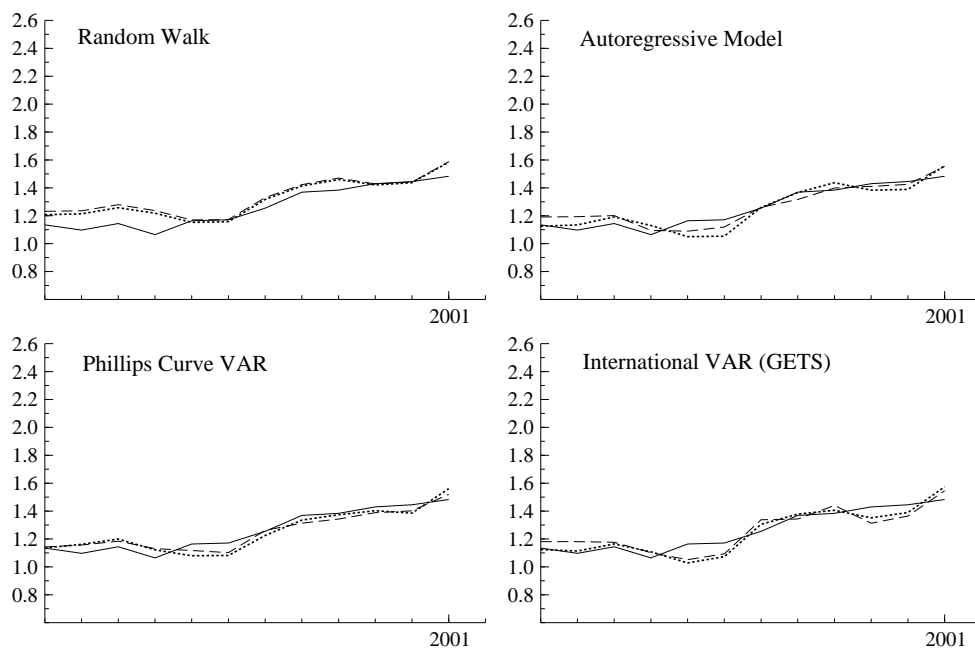


Figure 7: Year-on-year inflation rate of HICP excluding unprocessed food and energy and forecasts in %, 1 month ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

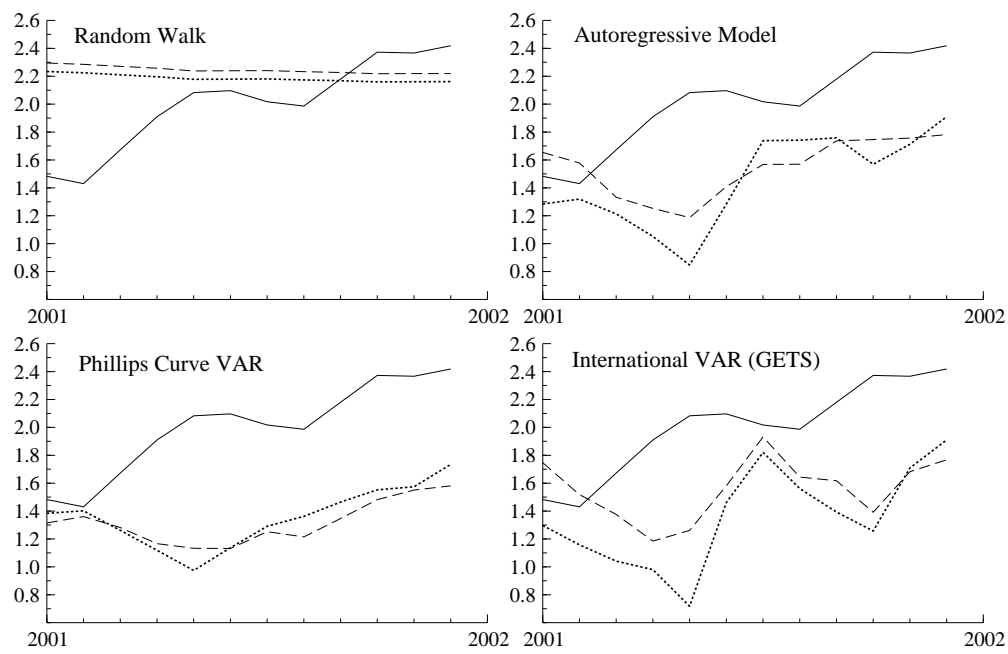


Figure 8: Year-on-year inflation rate of HICP excluding unprocessed food and energy and forecasts in %, 12 months ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

Here the results show a different pattern. All models exhibit a better accuracy for aggregating the subcomponent forecasts for a forecast horizon of one month except for the *RW* and the $VAR_{Gets}^{Int(2)}$. The $VAR^{Ph(12)}$ performs best overall. A similar pattern is found for the 12 months ahead forecasts except for the *RW* and $VAR^{Ph(12)}$. As in case of the overall year-on-year inflation, the *RW* performs best overall. However, Figures 7 and 8 shows that for the *RW* the difference between the indirect and direct approach to forecasting year-on-year inflation is negligible for both one month and 12 months horizons. The *AR* and the VAR_{Gets}^{int} models exhibit very similar forecasts of the direct and indirect approach for year-on-year 'core' inflation for one month ahead forecasts (Figure 7). In contrast, for 12 months ahead forecasts the difference is up to 0.6 percentage points (Figures 8), confirming the better RMSFE accuracy for the indirect method of aggregating subcomponent forecasts. For $VAR^{Ph(12)}$ both the direct and indirect approach perform similarly bad for the 12 months horizon as well as the 1 month horizon.

Overall, for the year-on-year inflation rate of HICP excluding unprocessed food and energy the RMSFE results indicate a better performance for the indirect method of aggregating the subcomponent forecasts, whereas the graphs with the actual forecasts show that in practice the improvement of the inflation forecast by aggregating subcomponent forecasts is not very large for some of the models considered in this forecast comparison. However, this 'core' inflation series including only those subindices of HICP that are less affected by unexpected shocks tends to be better forecasted by aggregating the subcomponent forecasts instead of forecasting the aggregate directly.

4 Conclusions: Why does disaggregation not necessarily help?

In this study an out-of-sample experiment is carried out to compare the relative forecast accuracy of aggregating the forecasts of euro area subcomponent inflation ('indirect' method) as opposed to forecasting aggregate euro area year-on-year inflation directly ('direct' method) in terms of their RMSFE. This study covers a broad range of models and model selection procedures. The results, although based on rather short samples, indicate that it is not necessarily better to aggregate the disaggregate component forecasts to forecast year-on-year aggregate euro area inflation.

In the light of these results the question arises why disaggregation does not necessarily help, especially at a 12 months horizon. Two arguments for aggregating disaggregated forecasts are the possibly better specification of the disaggregate equations implying a lower forecast error and that disaggregate forecast errors might cancel out when they are aggregated. On the other hand, some of the disaggregate components might be inherently difficult to model and to forecast. Therefore, it might be better to forecast the aggregate directly, since the aggregate series is usually smoother and easier to predict than some of the disaggregate components.

In the present study I analyse whether aggregating subcomponent forecasts delivers more accurate forecasts for aggregate year-on-year euro area inflation due to the larger information set used and better model specification than forecasting the aggregate directly. I find that a VAR model selected by a general-to-specific model selection procedure exhibits a lower forecast accuracy in many situations than a simple Phillips curve model. This finding indicates that a better in-sample model does not necessarily imply a lower forecast error out-of-sample. This result is in line with Clements & Hendry (1999, Ch3/4, 2001). Furthermore, one argument for the indirect approach to forecasting aggregate inflation, i.e. aggregating subcomponent forecasts, is also that this method allows for differences in the model specification across subcomponents. However, the analysis has shown that for the VAR where a general-to-specific model selection procedure selects from a relatively large number of potentially relevant domestic and international variables, it is better in terms of forecast accuracy to directly forecast aggregate year-on-year inflation for a forecast horizon of 12 periods ahead. The direct approach is favoured despite the varying specification across subcomponents resulting from the general-to-specific model selection procedure.

I also analyse the forecast errors for disaggregate components of euro area HICP and find that the forecast errors of the subcomponents do not cancel. This is because a shock, e.g. an oil price shock or a shock to unprocessed food in the euro area in 2001, affects several or even all components of HICP and therefore forecast errors appear in the same direction for those components affected. Therefore, the forecast bias of the aggregate is not reduced, but increased by aggregating the subcomponent forecasts in this case.

For inflation excluding unprocessed food and energy prices, sometimes referred to as 'core' inflation, the results are more favorable for aggregating subcomponent forecasts than in the analysis for overall HICP inflation. In that case the majority of models exhibits higher forecast accuracy for aggre-

gating subcomponent forecasts. Comparing these findings with the results for overall year-on-year inflation indicates that aggregating subcomponent forecasts is problematic when some subcomponents are inherently difficult to forecast due to frequent unexpected shocks to the series as in the case of energy and unprocessed food prices.

The findings suggest that to forecast euro area aggregate year-on-year inflation, aggregating subcomponent forecasts has to be considered with some caution since the resulting indirect aggregate forecasts might be less accurate than the direct forecast of aggregate inflation. For forecasting year-on-year inflation for the euro area the results presented so far raise the question whether modelling and forecasting the subcomponents is worthwhile if the forecast of the aggregate is the objective. One option might be to consider both the direct and the indirect method when forecasting year-on-year inflation for the euro area. There might be also a case for combining the two different forecast methods of forecasting the aggregate. Alternatively, the disaggregate models could be selected subject to the objective of forecasting the aggregate (for related proposals, see Van Garderen et al. (2000)). One interesting further line of research could explore the relative forecast accuracy of the direct versus the indirect method to forecasting inflation if such a model selection procedure would be employed.

To extend the range of models employed in the analysis, further research will also include dynamic factor models and Bayesian VARs that are recently more often employed in forecasting in a policy environment.

Appendix A Data

- to be completed -

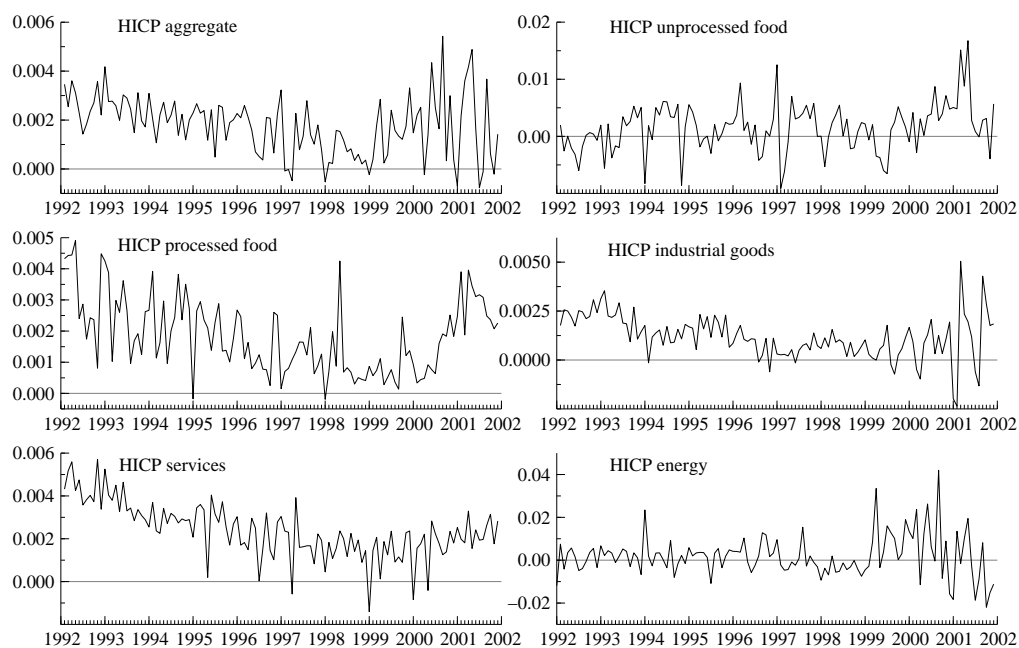


Figure 9: First differences of HICP (sub-)indices (in logarithm)

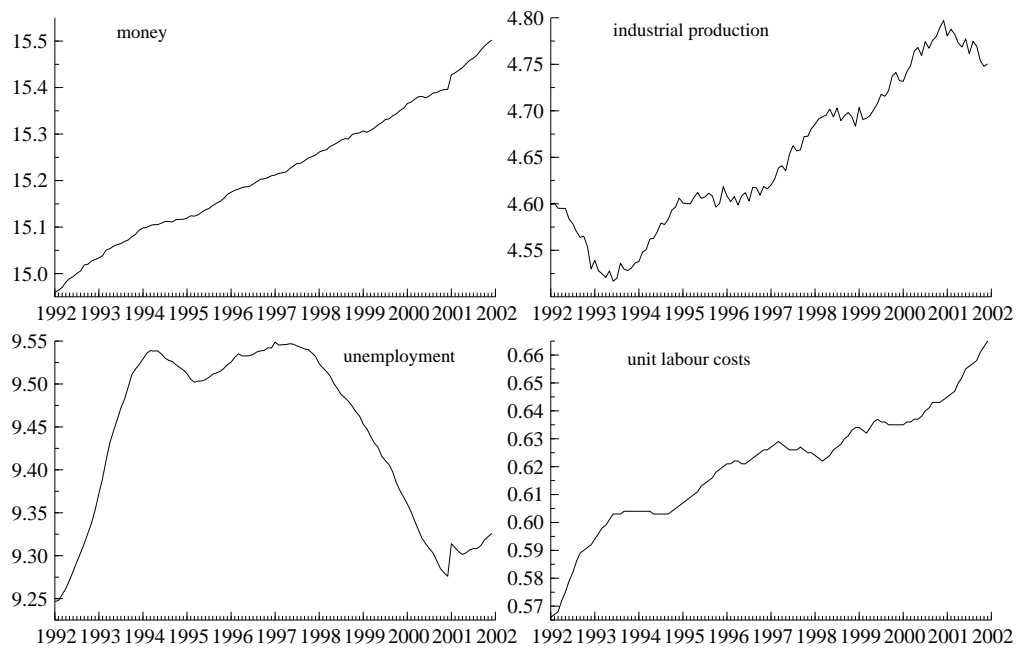


Figure 10: Money M3, industrial production, unemployment and unit labour costs (in logarithm)

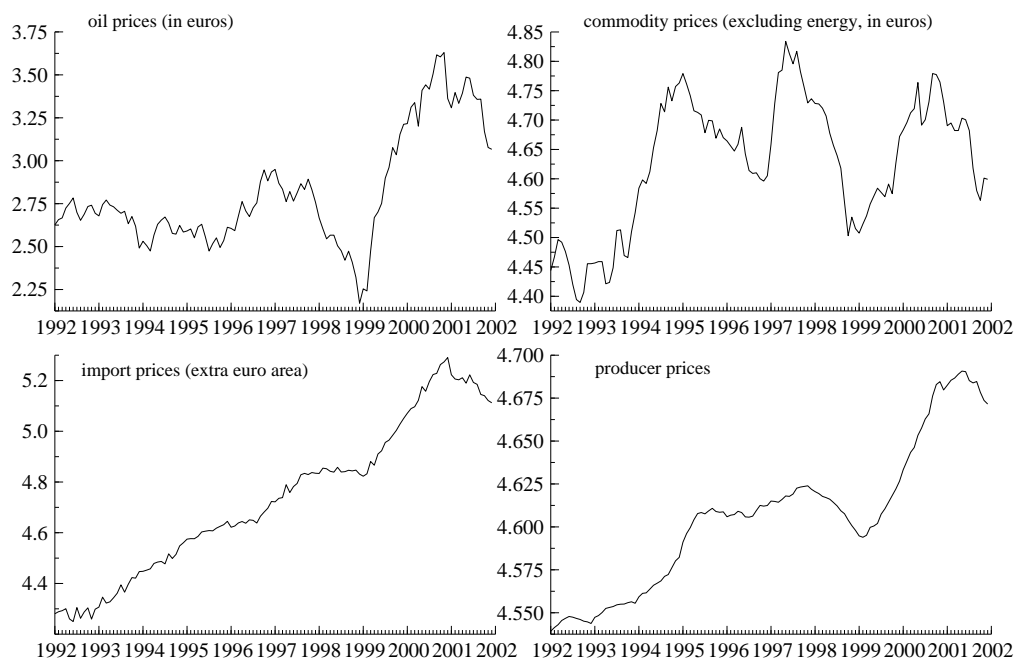


Figure 11: Oil prices (in euros), commodity prices (excluding energy, in euros), import prices (extra euro area), producer prices (in logarithm)

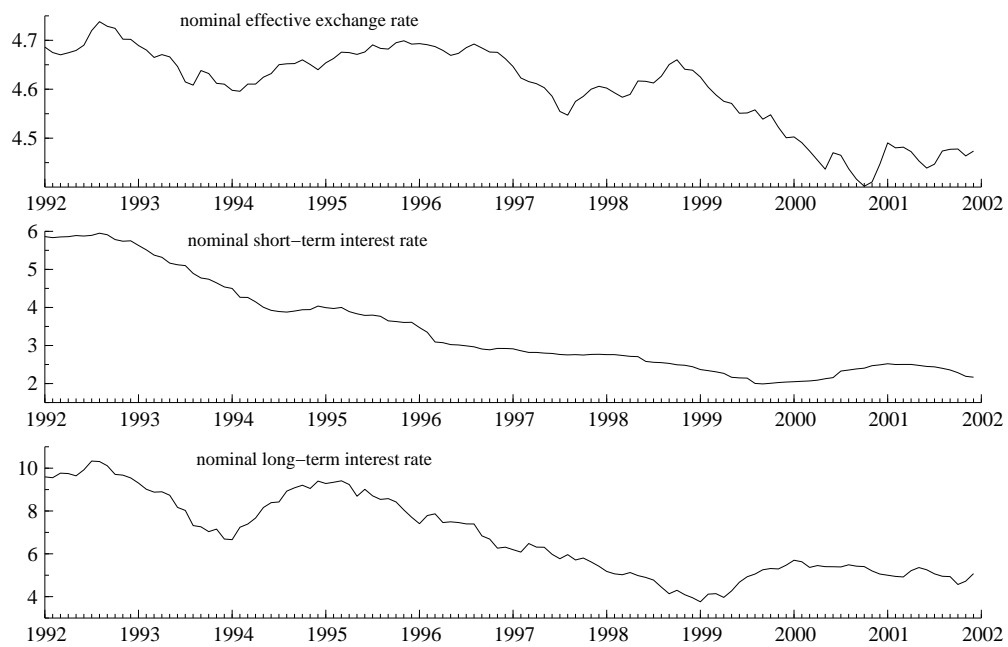


Figure 12: Nominal effective euro exchange rate, nominal short- and long-term interest rate (in logarithm except for interest rates)

References

- Angelini, E., Henry, J. & Mestre, R. (2001). Diffusion index-based inflation forecasts for the euro area, *ECB Working Paper 61*, European Central Bank.
- Barker, T. & Pesaran, M. H. (1990a). *Disaggregation in econometric modelling*, Routledge, London and New York.
- Barker, T. & Pesaran, M. H. (1990b). Disaggregation in econometric modelling - an introduction, in T. Barker & M. H. Pesaran (eds), *Disaggregation in econometric modelling*, Routledge, London and New York.
- Canova, F. (2002). G7 inflation forecasts, *ECB Working Paper 151*, European Central Bank.
- Clements, M. P. & Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*, MIT Press, Cambridge, Massachusetts.
- Clements, M. P. & Hendry, D. F. (2002a). Economic forecasting: Some lessons from recent research, *ECB Working Paper 82*, European Central Bank.
- Clements, M. P. & Hendry, D. F. (2002b). Modelling methodology and forecast failure, *Econometrics Journal* **5**: 319–344.
- Cristadoro, R., Forni, M., Reichlin, L. & Veronese, G. (2001). A core inflation index for the euro area, *Temì di Discussione 435*, Banca d'Italia.
- 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*, forthcoming.
- European Central Bank (2001). *The Monetary Policy of the ECB*.
- 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.
- Granger, C. W. J. & Morris, M. J. (1976). Time series modelling and interpretation, *Journal of Royal Statistical Society* **139**: 246–257.

- Granger, C. W. J. & Newbold, P. (1986). *Forecasting Economic Time Series*, 2nd edn, Academic Press, London.
- Grunfeld, Y. & Griliches, Z. (1960). Is aggregation necessarily bad?, *The Review of Economics and Statistics* **XLII**(1): 1–13.
- Hendry, D. F. & Krolzig, H.-M. (2001a). *Automatic Econometric Model Selection Using PcGets 1.0*, Timberlake Consultants Ltd., London, UK.
- Hendry, D. F. & Krolzig, H.-M. (2001b). New developments in automatic general-to-specific modelling, in B. P. Stigum (ed.), *Econometrics and the Philosophy of Economics*, forthcoming.
- Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past?, *Journal of Econometrics* (18): 337–349.
- Leamer, E. E. (1990). Optimal aggregation of linear net export systems, in T. Barker & M. H. Pesaran (eds), *Disaggregation in econometric modelling*, Routledge, London and New York, pp. 150–170.
- Lippi, M. & Forni, M. (1990). On the dynamic specification of aggregated models, in T. Barker & M. H. Pesaran (eds), *Disaggregation in econometric modelling*, Routledge, London and New York, pp. 150–170.
- Lütkepohl, H. (1984a). Forecasting contemporaneously aggregated vector ARMA processes, *Journal of Business & Economic Statistics* **2**(3): 201–214.
- Lütkepohl, H. (1984b). Linear transformations of vector ARMA processes, *Journal of Econometrics* (26): 283–293.
- Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag.
- Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis, *Journal of Business and Economic Statistics* **17**(1): 129–136.
- Marcellino, M., Stock, J. H. & Watson, M. W. (2002). Macroeconomic forecasting in the euro area: Country specific versus area-wide information, *European Economic Review*, forthcoming.

- Nijman, T. & Sentana, E. (1996). Marginalization and contemporaneous aggregation in multivariate garch processes, *Journal of Econometrics* **71**: 71–87.
- Rose, D. E. (1977). Forecasting aggregates of independent ARIMA processes, *Journal of Econometrics* (5): 323–345.
- 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. (1999). Forecasting inflation, *Journal of Monetary Economics* **44**: 293–335.
- Theil, H. (1954). *Linear Aggregation of Economic Relations*, North Holland, Amsterdam.
- Tiao, G. C. & Guttman, I. (1980). Forecasting contemporaneous aggregates of multiple time series, *Journal of Econometrics* (12): 219–230.
- Van Garderen, K. J., Lee, K. & Pesaran, M. H. (2000). Cross-sectional aggregation of non-linear models, *Journal of Econometrics* **95**: 285–331.
- Wei, W. W. S. & Abraham, B. (1981). Forecasting contemporaneous time series aggregates, *Communications in Statistics - Theory and Methods* (A10): 1335–1344.
- Zellner, A. & Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance, *Journal of Forecasting* **19**: 457–469.