EUROPEAN CENTRAL BANK

WORKING PAPER SERIES

WORKING PAPER NO. 247

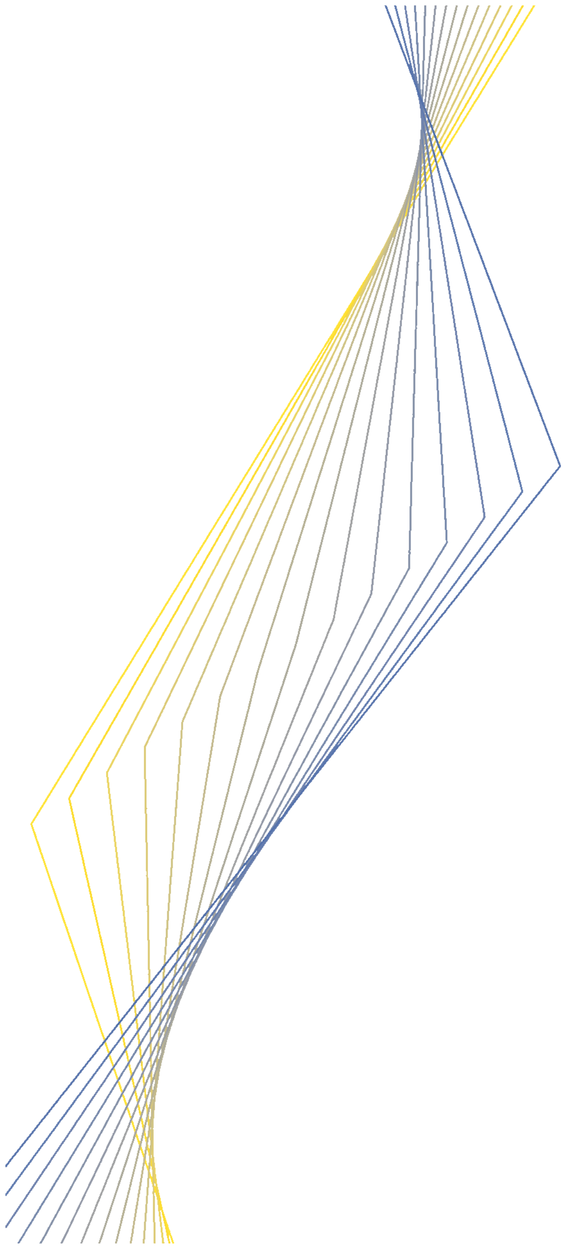
FORECASTING EURO AREA INFLATION: DOES AGGREGATING FORECASTS BY HICP COMPONENT IMPROVE FORECAST ACCURACY?

BY KIRSTIN HUBRICH

August 2003

EUROPEAN CENTRAL BANK

WORKING PAPER SERIES



WORKING PAPER NO. 247

FORECASTING EURO AREA INFLATION: DOES AGGREGATING FORECASTS BY HICP COMPONENT IMPROVE FORECAST ACCURACY?[1]

BY KIRSTIN HUBRICH[2]

August 2003

1 *A 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 Filippo Altissimo, Günter Coenen, Peter van Els, Neil Ericsson, Jan Groen, Tohmas Karlsson, Hans-Martin Krolzig, Helmut Lütkepohl, Massimiliano Marcellino, Jim Stock and an anonymous referee as well as participants of a seminar at the Dutch Central Bank, the European Econometric Society Meeting (ESEM) 2002 in Venice, a seminar of the Working Group of Econometric Modelling of the ESCB and a conference at the Banque de France for useful comments. The usual disclaimer applies. The opinions expressed herein are those of the author(s) and do not necessarily represent those of the European Central Bank. This paper can be downloaded without charge from http://www.ecb. int or from the Social Science Research Network electronic library at: http:// ssrn.com/abstract_id=457319.*

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

Address	**Kaiserstrasse 29**
	D-60311 Frankfurt am Main
	Germany
Postal address	**Postfach 16 03 19**
	D-60066 Frankfurt am Main
	Germany
Telephone	**+49 69 1344 0**
Internet	**http://www.ecb.int**
Fax	**+49 69 1344 6000**
Telex	**411 144 ecb d**

ISSN 1561-0810 (print)
ISSN 1725-2806 (online)

Contents

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 forecasting aggregate euro area inflation 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 forecast 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

Non-Technical Summary

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 forecasting aggregate euro area inflation can be improved by aggregating forecasts of subindices of the Harmonized Index of Consumer Prices (HICP) as opposed to forecasting the aggregate HICP directly. A simulated out-of-sample forecast 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 root mean square forecast error. This study covers a broad range of models and model selection procedures, including univariate and multivariate linear time series methods, and it 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. Some of the methods employed imply the same specification across all subcomponents and the aggregate, whereas others allow for different lag lengths and different macroeconomic variables to be included. This set-up allows to investigate to what extent different model specifications influence the forecast results.

The results indicate that aggregating forecasts by component does not necessarily help forecast year-on-year inflation twelve months ahead. For forecasting year-on-year inflation in the euro area the results presented raise the question whether modelling and forecasting the subcomponents is worthwhile if the forecast of the aggregate is the objective.

Although the details of the results in this study are of course specific to the empirical application of euro area inflation, the findings nevertheless point at some more general problems the forecaster may face when aggregating forecasts of disaggregate components to forecast the aggregate. The following main explanations for the results can be derived. The results suggest that taking into account differences in the dynamic properties of subcomponents by different model specifications across subcomponents in terms of variables and/or lags do not necessarily improve the aggregate forecast. Even for a forecast method where the variables that enter the respective model for each of the HICP subcomponents are selected 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 months ahead. Furthermore, combination of different forecast methods as well as the direct and indirect forecasts is found not to improve over the (best) direct forecast 12 months ahead.

The forecast errors for disaggregate components of euro area HICP are also analysed. It turns out that the forecast errors of the subcomponents do not cancel. This is because many shocks, e.g. the oil price shock or the shock to unprocessed food in 2000 and 2001 in the euro area, affect several or even all components of HICP over the evaluation period and therefore forecast errors appear in the same direction for those components affected. Also, the extent of the effect of a shock and its dissemination across subcomponents is difficult to predict. Therefore, the forecast bias of the aggregate is not reduced, but increased by aggregating the subcomponent forecasts in this case.

Furthermore, the forecast performance of aggregating subcomponent forecasts is investigated for another inflation measure of interest to monetary policy makers: 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. For this aggregate the majority of methods exhibits higher forecast accuracy for aggregating subcomponent forecasts. Comparing these findings with the results for overall year-on-year inflation leads to further insight into the reasons for the problems with aggregating disaggregate forecasts: Aggregating subcomponent forecasts appears to be problematic when some subcomponents are inherently difficult to forecast due to frequent shocks to the series, in case of HICP the subcomponents energy and unprocessed food prices.

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, according to the clarification in May 2003, as a year-on-year increase in the Harmonized Index of Consumer Prices (HICP) for the euro area of below, but "close to 2% over the medium term" (European Central Bank, 2003b).

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 subcomponents of inflation and aggregating forecasts based on these models. Contemporaneous aggregation of forecasts may be considered in two dimensions: the aggregation of national HICP forecasts for euro area countries and the aggregation of HICP subcomponent forecasts for the euro area.

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 aggregate HICP directly as opposed to aggregating forecasts for HICP subcomponents. A broad range of models and model selection procedures is employed. The comparison is based on data for the euro area as a whole as these are the data relevant for the monetary policy of the ECB.

The debate about aggregation versus disaggregation in economic modelling goes back to Theil (1954) and Grunfeld & Griliches (1960). One strand of that literature has focussed on the effect of contemporaneous aggregation on forecast accuracy. There are two main arguments for aggregating forecasts of disaggregated variables instead of forecasting the aggregate variable of interest directly. One rationale is that the disaggregated variables can be better modelled by taking their different dynamic properties into account

[3] The 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.

and, therefore, can be predicted more accurately than the aggregate variable. Modelling disaggregated variables may involve using a larger and more heterogenous information set, and specifications may vary across the disaggregate variables (see Barker & Pesaran, 1990b). A second argument in favour of disaggregation is that forecast errors of disaggregated components might cancel partly, leading to more accurate predictions of the aggregate (see also Clements & Hendry (2002b) for a discussion on forecast combination as bias correction). In contrast, it may also be argued that it is 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 shocks to some of the disaggregate variables, as will be seen in the analysis presented in this study. On the other hand, a well specified model does not necessarily imply higher forecast accuracy. An additional argument against disaggregation for forecasting the aggregate is that 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 analyse the role of a number of factors that based on asymptotic theory and Monte Carlo simulations have been found in the literature to affect the role of disaggregation on forecasting accuracy. They 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 versus 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, 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)).[4] Non-linear time series models are not considered here due to the short time series available for estimation.[5] Time-varying parameter models are not considered here, either. Stock & Watson (1996) suggest that gains

[4]For an application of a VECM to forecasting euro area inflation taking into account a cointegration relation between HICP subindices, see Espasa, Senra & Albacete (2002).

[5]An exposition of the forecasting performance of non-linear models can, for example, be found in Clements & Hendry (1999, Ch.10). See also Marcellino (2002) for some promising results using non-linear models for longer euro area macroeconomic series that are extended backwards by aggregating available country data.

from using time varying parameter models for forecasting are generally small or non-existent, especially for short horizons.[6] Misspecification is not a central issue in the context of the present paper. For example, Diebold & Kilian (2000) show that a correctly specified model does not necessarily improve the forecast accuracy relative to a misspecified model.

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. The model selection procedures also include the Schwarz information criterion for selecting parsimonious models as well as a general-to-specific modelling strategy implemented in the software package PcGets (Hendry & Krolzig, 2001a).[7] The latter model selection procedures allow for varying specifications across subcomponents in terms of lag order and / or variables included.

The remainder of the paper is structured as follows: In section 2 some asymptotic and small sample simulation results from the literature regarding the relative forecasting performance of aggregated forecasts of time series subcomponents are discussed. Section 3 presents the data used in the analysis. The forecast methods and model selection procedures employed on which the forecast comparison is based are outlined in section 4, whereas in section 5 the empirical results for the relative forecast accuracy of the aggregated versus the disaggregated approach to forecasting euro area inflation are presented and discussed. Finally, section 6 draws some tentative conclusions from the analysis.

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.[8] The focus of this study is on analysing the effects of contemporaneous aggregation of subcomponents of

[6]See Canova (2002) for a recent more favourable evaluation of the forecasting performance of Bayesian time varying parameter models, whereas his results for BVARs are less favourable.

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

[8]In 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.

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

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

where y_t^j $(j = 1, ..., n)$ are the subcomponents of y_t^{agg}, n is the number of subcomponents considered and w_j, $j = 1, ..., n$, are the aggregation weights. It is assumed that the aggregation weights are fixed, i.e. they do not change over time[9], and that $w_j > 0 and \sum w_j = 1$. 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 $\hat{y}_t^j$ $(j = 1, ..., n)$, i.e. $\hat{y}_{sub,t}^{agg} = \sum w_j \hat{y}_t^j$.

The issue of contemporaneous aggregation of economic variables has already been discussed and analysed in an early contribution by Theil (1954) who argues that a disaggregated modelling approach improves the model specification of the aggregate. Grunfeld & Griliches (1960), however, point out that if the micro equations are not assumed 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 focused 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.[10]

In the following, the main asymptotic and small sample simulation results from the latter strand of the literature, that are of interest in the

[9]For an extension allowing for time varying weights, see e.g. Van Garderen, Lee & Pesaran (2000).

[10]Some 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).

context of the empirical analysis presented later, are summarized. Based on asymptotic theory the following results can be derived. If the DGPs of the individual subcomponents and the aggregate are known in terms of structure and coefficients, aggregating subcomponent forecasts is better in terms of a mean square forecast error (MSFE) criterion than forecasting the aggregate directly, MSFE($\hat{y}^{agg}_{sub}$) < MSFE($\hat{y}^{agg}$). This result is due to the larger information set underlying the aggregate forecast. However, the usefulness of this result is limited because in practice the DGP is usually not known. If the assumption of a known DGP is relaxed and it is assumed that the unknown process order is estimated using a consistent order selection criterion, the relative forecast accuracy of the direct or indirect approach to forecasting the aggregate will depend on the true DGP.[11] Under certain assumptions about the DGP[12] the aggregation of forecasts of the components can actually be inferior to forecasting the aggregated time series directly, MSFE($\hat{y}^{agg}_{sub}$) > MSFE($\hat{y}^{agg}$). The higher estimation variability of estimating the disaggregated processes instead of the aggregate process may increase the MSFE, in some cases even in large samples (Lütkepohl, 1987, p.310). The relative forecast accuracy depends on the extent to which the systematic differences in the MSFE are offset by the effects of estimation variability.

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 subcomponent forecasts. He also includes modelling approaches where parsimonious specification is limiting estimation variability due to reduced precision of the estimates in short samples. 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

[11] In case of a finite order DGP the asymptotic MSFE matrices are the same as in the case of known process orders. In the case of an infinite order DGP an approximation of the MSFE can be derived asymptotically under the assumption that the AR orders of the processes fitted to the data approach infinity with the sample size (Lütkepohl, 1987, p.73).

[12] See Lütkepohl (1987, p.129)

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. The aim is to gather insights about the effect of contemporaneous aggregation on forecasting accuracy for euro area HICP inflation at the center of interest of the ECB's monetary policy.

I extend earlier studies in paying special attention to the potential role of macroeconomic predictors for HICP inflation. The theoretical literature and the simulation studies on the effects of contemporaneous aggregation on forecasting accuracy mainly focus on univariate models. Only those multivariate models are considered that include different HICP subcomponents, but no other macroeconomic variables are introduced (see e.g. Lütkepohl, 1987).

Since in this study I investigate the role of macroeconomic variables in forecasting aggregate HICP inflation versus forecasting subcomponent HICP inflation, I employ different model selection procedures. In contrast, most of the asymptotic results regarding (dis-)aggregation in forecasting 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 forecast efficiency of models with a high number of parameters. Therefore, the model selection procedure is important in deriving the final model. One possibility is to employ information criteria to select the lag length of (V)AR models. Lütkepohl (1984a) presents results for subset VARs where information criteria are also used to decide on deleting individual elements from the coefficient matrices. This model selection procedure leads 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 (for more details on the model selection procedure, see 4).

3 HICP aggregate data and subcomponents

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

This particular breakdown into subcomponents has been chosen in accordance with the data published in the ECB Monthly Bulletin and since the analysis of price developments of HICP subcomponents regularly presented in the ECB Monthly Bulletin (see European Central Bank (2000, p.28)) is based on this breakdown. A range of explanatory variables for inflation is also considered.

The data employed are of monthly frequency[13], starting in 1992(1) until 2001(12). This is a relatively short sample, which is determined by the availability of data for the euro area. The sample is split into an estimation and a forecast evaluation period. Model selection and estimation is carried out on the basis of 36 recursive samples starting from 1992(1) up to 1998(1), extending the sample by one month sequentially. The longest recursive estimation sample ends in 2000(12). Seasonally adjusted data have been chosen[14] because of the changing seasonal pattern in some of the HICP subcomponents for some countries due to a measurement change.[15],[16] The notation for the HICP subindices 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}.

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 respective indices are depicted in Figure 2. The HICP indices in first differences are displayed in Figure 8 in the Appendix. Aggregate HICP, HICP processed food, HICP industrial production and HICP services in levels 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 aggregate 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 deseases.

Since Diebold & Kilian (2000) show for univariate models that testing for a unit root is useful for selecting forecasting models, Augmented Dickey Fuller (ADF) tests have been carried out for all HICP (sub-)indices (in logarithm). The tests are based on the sample from 1992(1) to 2000(12), i.e. the longest of the recursively estimated samples. The tests do not reject non-stationarity for the levels of all (sub-)indices over the whole period.[17]

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

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

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

[16]The sensitivity of the results to using seasonally unadjusted data has been analysed on the basis of a shorter sample. The results show no substantial change in the conclusions.

[17]The ADF test specification includes a constant and a linear trend for the levels and first differences. The number of lags included is chosen according to the largest significant

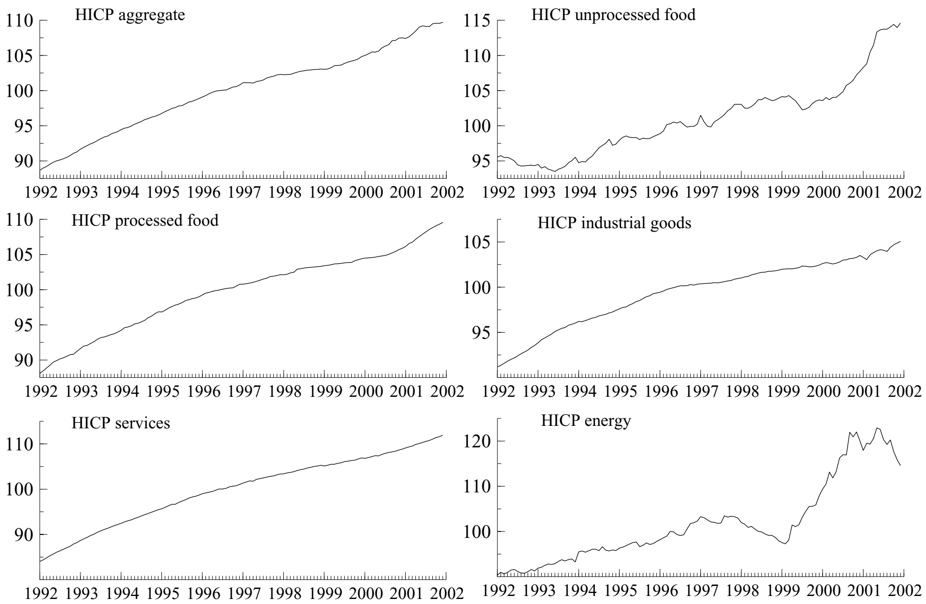


Figure 1: HICP aggregate and subindices (in logarithms)

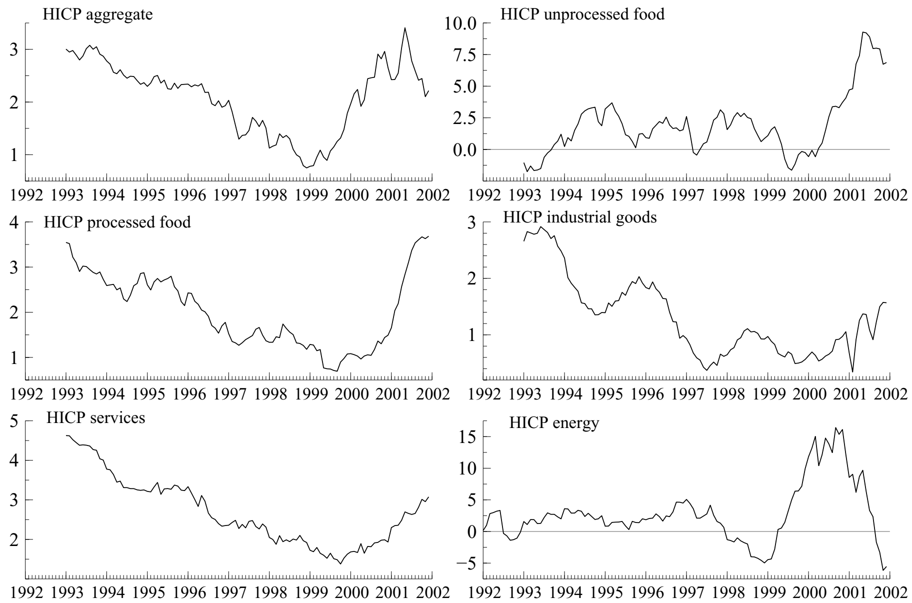


Figure 2: Year-on-year HICP inflation (in %), aggregate and subindices

Non-stationarity is rejected for the first differences of all series except the aggregate HICP and HICP services. For the first differences of the latter two series, however, non-stationarity is not rejected for all shorter recursive estimation samples up to 2000(8) and 2000(7), respectively. Therefore and because of the low power of the ADF test HICP (sub-)indices are assumed to be integrated of order one in the analysis and modelled accordingly.

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 (extra euro area), 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, $NEER$[18], 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.[19]

4 Forecast Methods and Model Selection

Five different forecasting models using different model selection procedures are employed for both forecast methods, i.e. forecasting HICP inflation directly and aggregating subcomponent forecasts. In case of the first three forecasting models the specification is the same across HICP subcomponents. The random walk with drift (*RW*) is employed as a benchmark model since it has often been found to outperform other forecasting models. Furthermore, a simple Phillips curve model, as e.g. in Stock & Watson (1999b), is employed including inflation and the change in unemployment in the VAR with 12 lags. This model will be denoted $VAR^{Ph(12)}$. The third model is a large VAR with 12 endogenous domestic and international variables described in the data section above, allowing for 2 lags only due to the short sample ($VAR^{Int(2)}$). The fourth and fifth model are chosen based on in-sample information. A univariate autoregressive (AR) model is included in the comparison where the lag order is parsimoniously chosen using the Schwarz criterion, denoted AR^{SC}. Therefore, the lag order varies across the different components. Finally, a general-to-specific model selection strategy is employed to choose a VAR (VAR^{Int}_{Gets}), implemented in the computer package PcGets by Hendry & Krolzig (2001a), where the choice of variables

lag on a 5% significance level.

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

[19]A reliable measure of administered prices and indirect taxes to be included in the analysis is not available for the euro area (European Central Bank, 2003a).

and lag length is based on mis-specification tests, structural break tests, t- and F-block tests, encompassing tests and information criteria. A 'liberal' selection strategy has been chosen implying a higher probability of retaining relevant variables at the risk of retaining irrelevant ones. Since PcGets is in principle a single equation procedure, a LR test has been carried out to test the null hypothesis that the specific models selected by PcGets for each of the variables included in the VAR are a valid reduction of the unrestricted redued form VAR. This test does not reject for the models employed in the analysis. The model is selected by PcGets starting with a VAR including the large potential number of domestic and international variables as included in $VAR^{int(2)}$. In contrast to $VAR^{int(2)}$, for this model type different variables and lag lengths are possibly chosen across different HICP subcomponents and the aggregate. The two methods AR^{SC} and VAR^{Int}_{Gets} are included to analyse whether different specifications across subcomponents in terms of lags and variables help to improve the forecasting accuracy of aggregating subcomponent forecasts. The automated model selection procedure implemented in PcGets is particularly useful in this investigation since economic theory does not provide much guidance on how to model the disaggregate components of HICP. It should be noted, however, that the general-to-specific model selection procedure implemented in PcGets does not aim at improving forecast performance, but is based purely on in-sample information.

Simulated out-of-sample model choice has not been considered here because it is not feasible nor is it desirable. It is not feasible, since the already short sample would have to be split further to create an additional out-of-sample period for model choice different from the forecast evaluation period. It is not desirable either to choose the models based on a simulated out-of-sample experiment, since the model choice would depend on the specific characteristics of the out-of-sample model selection period and whether the DGP in this short period resembles the DGP in the future. Also, Inoue & Kilian (2003) show that the simulated out-of-sample selection method will select overparameterised models with positive probability, resulting in larger finite sample RMSFE.

For all models, except for VAR^{Int}_{Gets} where PcGets is employed, the model choice and simulated out-of-sample forecast experiment are carried out using GAUSS. All models are re-estimated for each of the recursive samples. Regarding the model selection procedures, the AR^{SC} is applied for each of the recursive samples. The lag lengths for the different component models and the aggregate model do hardly change over the different recursive samples, however. The PcGets procedure is applied for the sample until 1998(1).[20]

[20] For a shorter forecast evaluation period, namely over the last 12 out-of-sample periods, PcGets has been applied to choose a new model for each recursive sample. This did not improve the relative performance in comparison with the other methods.

5 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 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 forecast comparison between different representations of the same system. Furthermore, switching the basis of comparison can lead to a change in ranking of the methods in this case.[21] 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 aggregate HICP is a weighted chain index, where the weights change each year. Since the end of all recursive estimation samples is in 1998, 1999 and 2000, respectively, the aggregation of the forecasts is carried out using the HICP subcomponent weights of the respective end year of the estimation period (at prices of December the previous year) which would be known to the forecaster in real time.[22] The forecasts from the models in first differences are recalculated to level forecasts and rebased to the month 1997(12), 1998(12) and 1999(12), respectively, 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 and transformed into year-on-year inflation rates. Those are then compared with the respective realization of year-on-year inflation. The actual weights used of, for example, the year 2000 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.

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 indirect forecast of aggregate inflation, i.e. the aggregated forecasts of the subindices ($\Delta_{12}\hat{p}^{agg}_{sub}$).

Tests of equal forecast accuracy or forecast encompassing tests[23] have

[21] Clements & Hendry (1998, p.69/70).

[22] Note that therefore one source of the resulting forecast error is also the change in subcomponent weights in the following year in comparison to the current year, although the changes in weights from year to year are relatively small.

[23] Taking into account estimation uncertainty West & McCracken (1998) and West (2001) propose tests of equal forecast accuracy and forecast encompassing for non-nested models. In contrast, Clark & McCracken (2001) present forecast accuracy and forecast encompassing tests for nested models.

not been carried out, since we compare methods rather than models, i.e. forecasting the aggregate directly is compared to aggregating the subcomponent forecasts. This distinction between methods and models is in line with Stock & Watson (1999a, p.2), who refer to forecast methods in the case where forecasts are based "not on a single estimated model but on results from multiple models that are estimated subject to model selection criteria or pretests."[24]

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. In the paper, the results for 1-,6- and 12-months ahead forecasts are presented. The RMSFE evaluation is based on recursive forecasts that involve an average of the respective horizon forecasts over all 36 recursive samples.[25] The one step ahead forecasts are starting with the forecast for 1998(2) based on the estimation sample 1992(1) to 1998(1), the second forecast is for 1998(3) based on the estimation sample up to 1998(2), etc., the 36th forecast for 2001(1) is then based on the estimation sample up to 2000(12). Similarly, 12-period-ahead forecasts are carried out for 36 different estimation samples. The forecast for 1999(1) is based on the sample up to 1998(1), whereas the last 12 step ahead forecast is carried out for 2001(12) based on the estimation sample until 2000(12).

Other simulated out-of-sample experiments have been carried out considering 3 subperiods of 12 months of the forecast evaluation period to analyse the sensitivity of the results towards a specific forecast period. The results of this analysis did not change the conclusions of the paper.[26] The following presentation of the forecast comparison focusses on the longest forecast evaluation 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 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}_{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 in unprocessed food prices in 2000 by explicitly including

[24]A similar argument is brought forward by Stock & Watson (2003). In the context of comparing direct and aggregated forecasts across euro area countries Marcellino et al. (2002) is an example for presenting no tests of forecast accuracy nor of forecast encompassing, focusing on MFE and MSFE results.

[25]Note that in this paper due to the short estimation and forecast evaluation period, the forecast origins are kept the same for all forecast horizons. Additional forecasts for shorter horizons at a different forecast origin implying different parameter estimates might in this case have a comparatively large impact on the average performance of the different forecast methods.

[26]The results are available from the author upon request.

Table 1: **Relative forecast accuracy, RMSFE of year-on-year inflation in percentage points, Recursive estimation samples 1992(1) to 1998(1),...,2000(12)**

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$
RW	0.142	0.142	0.476	0.479	0.807	**0.813**
$VAR^{Ph(12)}$	0.146	0.154	0.456	0.444	1.063	0.980
$VAR^{Int(2)}$	**0.115**	**0.103**	0.422	**0.437**	0.788	**0.814**
AR^{SC}	0.144	0.143	0.439	0.475	0.756	0.877
VAR^{Int}_{Gets}	0.130	0.111	**0.404**	0.449	**0.702**	0.842

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest RMSFE per column

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 VAR^{Int}_{Gets} turns out to be best overall for the period considered for $h = 6$ and $h = 12$ for directly forecasting the aggregate. The $VAR^{Int(2)}$ model performs better than most of the other models for $h = 6$ and 12 for the indirect method. This indicates that a different specification across components chosen based on in-sample criteria does not necessarily improve the forecast accuracy of aggregating subcomponent forecasts. The average RMSFE and average mean forecast error (MFE) for all forecast horizons is presented in the Table 3. These numbers also exhibit higher forecast accuracy of the direct forecast method for most models on average over the forecast horizons.

The MFE in Table 2 shows that the modulus of the bias of the forecast tends to be lower for some of those methods that also show a lower RMSFE in the case of aggregating the subcomponent forecasts ($\Delta_{12}\hat{p}^{agg}_{sub}$) for a one month horizon. For 6 and 12 months forecast horizons, in contrast, the direct forecast method exhibits a lower MFE for most models, as in Table 1 in terms of RMSFE.

It should be noted that the general-to-specific model selection procedure for VAR^{Int}_{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 a twelve months ahead forecast, the VAR^{Int}_{Gets} does improve forecast accuracy in RMSFE terms over the $VAR^{Int(2)}$ for forecasting the aggregate directly, but not for aggregating the subcomponent forecasts where the $VAR^{Int(2)}$

Table 2: **Relative forecast accuracy, MFE of year-on-year inflation in percentage points, Recursive estimation samples 1992(1) to 1998(1),...,2000(12)**

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$
RW	-0.044	0.050	-0.161	0.191	-0.233	**0.295**
$VAR^{Ph(12)}$	0.022	**0.0**	0.205	-0.170	0.488	-0.432
$VAR^{Int(2)}$	**0.001**	-0.004	0.115	**-0.132**	0.304	-0.334
AR^{SC}	-0.027	-0.012	-0.110	-0.156	-0.141	-0.378
VAR^{Int}_{Gets}	-0.027	-0.008	**-0.083**	-0.155	**-0.118**	-0.359

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest MFE (in absolute terms) per column

Table 3: **Average relative forecast accuracy over all forecast horizons, average RMSFE and average MFE of year-on-year inflation in percentage points, Recursive estimation samples 1992(1) to 1998(1),...,2000(12)**

criterion	ARMSFE		AMFE	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$
RW	0.494	0.497	-0.162	0.195
$VAR^{Ph(12)}$	0.535	0.519	0.232	-0.193
$VAR^{Int(2)}$	0.443	**0.452**	0.135	**-0.152**
AR^{SC}	0.460	0.507	-0.107	-0.177
VAR^{Int}_{Gets}	**0.421**	0.471	**-0.085**	-0.172

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest ARMSFE and AMFE per column surprisingly performs better. This indicates that the model selection procedure that is designed to improve the in-sample fit of the respective method, does not necessarily improve the out-of-sample forecast accuracy for the HICP subcomponents (see also Table 4).[27]

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

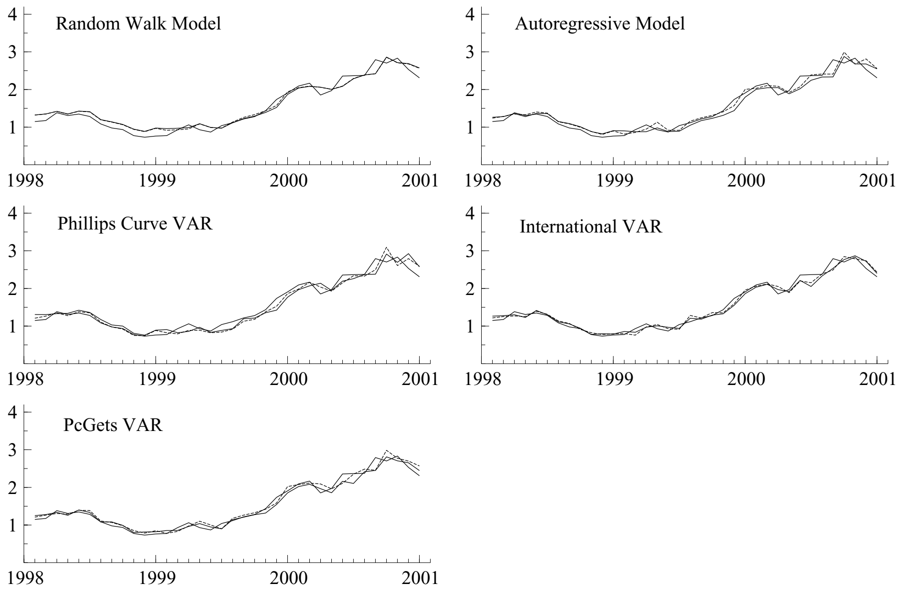


Figure 3: Year-on-year inflation rate and forecasts in %, 1 month ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

Figure 3 presents the actual year-on-year inflation rates and the 1-step ahead forecasts for 36 recursive samples for all methods. 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 for any of those models. 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 the RW model there is hardly any difference between the direct and indirect fore-

[27]The relation between model (mis-)specification and forecast accuracy has been discussed extensively in e.g. Clements & Hendry (1999, Ch3/4, 2001); see also Clements & Hendry (2002a).

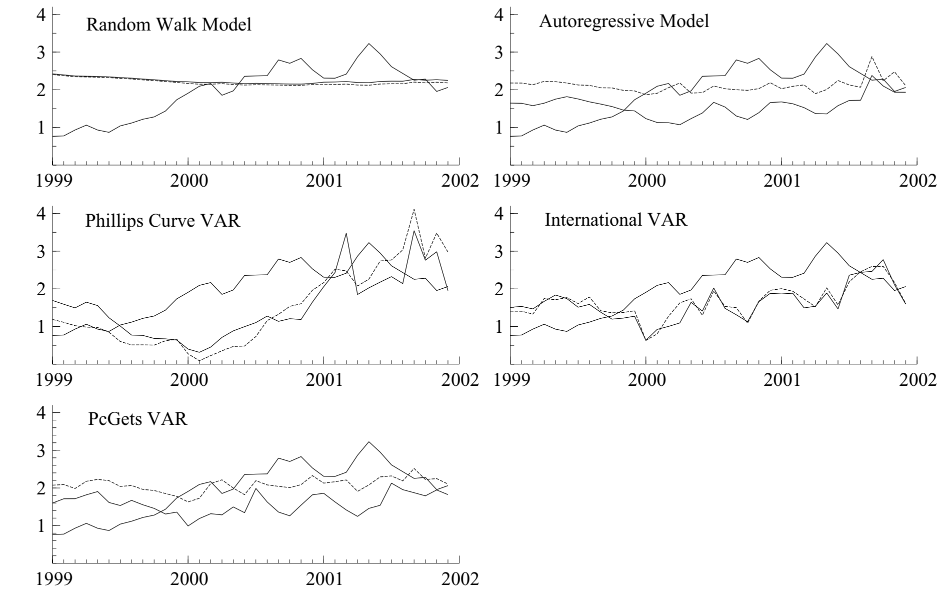


Figure 4: Year-on-year inflation rate and forecasts in %, 12 months ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

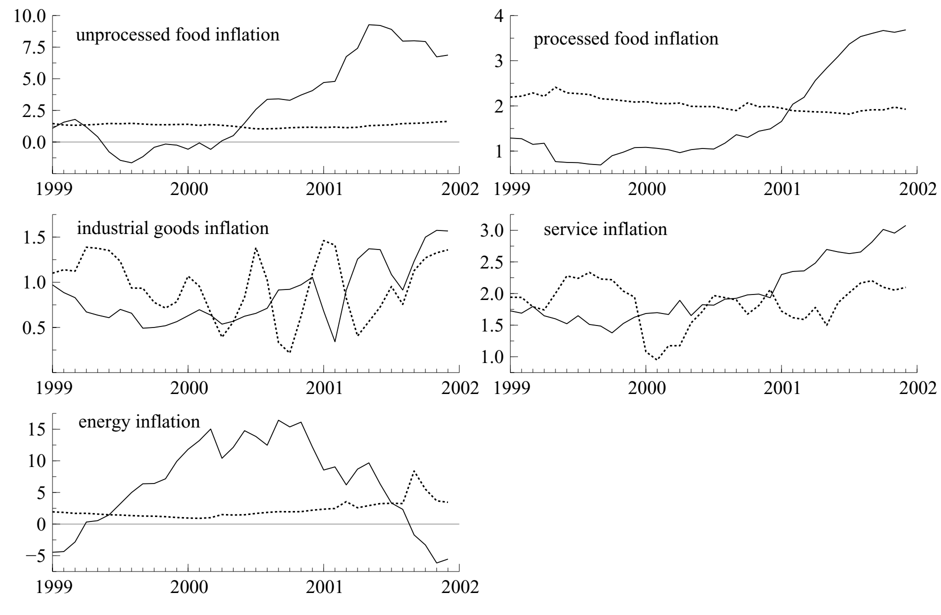


Figure 5: Year-on-year inflation rate in %, solid: actual, dashed: AR subcomponent forecast, 12 months ahead

cast, whereas the AR^{SC}, the $VAR^{Ph(12)}$, the $VAR^{Int(2)}$ and the VAR^{Int}_{Gets} differ up to around 1.3 percentage points. For the majority of those models that exhibit a relevant difference between the direct and indirect forecast of year-on-year inflation, i.e. AR^{SC}, $VAR^{Int(2)}$ and VAR^{Int}_{Gets}, the RMSFE indicates a better performance of forecasting the aggregate year-on-year inflation directly. The predictive failure of all methods for the 12 months ahead forecast over most of the recursive samples can be explained by their failure to predict several unexpected events: The increase in year-on-year changes of unprocessed food prices since early 2000 due to the effects of weather conditions and 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 in 2001, which is to a large extent due to lagged effects of the increase of energy prices and the depreciation of the euro. Furthermore, the increase in year-on-year changes of energy prices since 1999 and its decline in 2001 is not well captured by either of the methods.

Figure 5 shows the results for the AR model for each of the subcomponents since this model performs comparatively well. It can be seen that unprocessed food, processed food and services inflation are over-predicted in the beginning of the forecast evaluation period, whereas especially unprocessed food and processed food inflation are substantially under-predicted for the whole year of 2001. Energy inflation is substantially over-predicted for the second half of 1999, the year 2000 and the first half of 2001. A similar picture arises for the other models. All forecast models fail badly in predicting the most volatile HICP components, p^{uf} and p^{e}. Table 4 presents the respective RMSFE per component. These results provide some explanation why aggregating subcomponent forecasts is not better than forecasting the aggregate inflation rate directly: The subcomponents are affected by certain shocks in the same way and therefore lead to forecast failures in the same direction.

Overall, for aggregate HICP the results presented in this study reveal a tendency to higher forecast accuracy of forecasting aggregate year-on-year inflation directly over longer horizons, especially for the 12 months horizon of interest for monetary policy.

Forecast Combination One line of further interesting research is whether combining direct and indirect forecasts improves forecast accuracy over all the forecasts of the aggregate 12 months ahead. Two possible directions of forecast combination are of interest here. Either combining the direct and indirect forecast for each of the five methods considered or, alternatively, combining the different direct forecasts based on alternative methods and compare those forecasts to combined indirect forecasts from different methods. A comprehensive discussion of different methods of forecast combina-

Table 4: **Forecasts of HICP (sub-) indices: Forecasting accuracy, RMSFE of year-on-year inflation, forecast horizons: 1 and 12**

	RW	$VAR^{Ph(12)}$	$VAR^{Int(2)}$	AR^{SC}	VAR^{Int}_{Gets}
h=1					
p^{uf}	0.365	0.502	0.466	**0.361**	0.393
p^{pf}	0.119	0.106	0.108	**0.098**	0.111
p^{i}	0.106	**0.098**	0.115	0.100	0.110
p^{e}	1.442	1.756	1.169	1.493	**1.168**
p^{s}	0.149	0.104	0.122	**0.097**	0.113
h=12					
p^{uf}	3.783	4.316	3.998	**3.677**	3.722
p^{pf}	1.233	1.229	**0.962**	1.134	1.127
p^{i}	0.815	0.570	0.534	**0.473**	0.594
p^{e}	7.957	12.477	8.622	8.196	**7.775**
p^{s}	1.409	0.604	0.592	0.532	**0.500**

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest RMSFE per component row tion would go beyond the scope of the paper. A discussion of the different forecast combination methods and review of the literature can be found in Clemen (1989), Diebold & Lopez (1996) and Clements & Hendry (2002c).

It should be noted that in principle the aggregation of subcomponent forecasts is a way of forecast combination. As discussed above, combining the subcomponent forecasts does not help to forecast the aggregate 12 months ahead since forecasts will fail in the same direction when an unexpected shock occurs that is affecting some or all forecasts to be combined. On the same grounds, one would expect that combining the direct and indirect method of forecasting aggregate inflation does not necessarily improve forecast accuracy. On the other hand, combination of forecasts can improve the overall forecasts if models provide partial explanations, especially if forecasts are differentially biased (one is biased upward, one downward). Furthermore, variance reduction can be achieved by using various information sets efficiently. Sample estimation uncertainty will also influence the relative forecast accuracy. Clements & Hendry (2002c) derive some results why forecast combination might work for one-step ahead forecasts. However, there are to the knowledge of the author no results available for multi-step ahead forecasts. Therefore, whether forecast combination improves over separate forecasts of aggregate inflation has to be investigated empirically.

In the context of this study, RMSFEs are calculated for a number of combined forecasts of euro area inflation 12 months ahead. Simple (mean) averaging is employed since that is often found to perform better than more sophisticated methods (see e.g. Clements & Hendry, 2002c). For three out of the five methods considered the direct and the indirect forecasts are actually biased in the opposite direction (see Table 2). Thus, the direct and indirect forecast for each method are combined. However, forecast combination does not improve the RMSFE over the best forecast for the respective method. Alternatively, different forecast methods are combined for the indirect forecast and compared with different combined forecast methods for the direct forecast. More precisely, since for the indirect forecast the RW and the $VAR^{int(2)}$ perform best in RMSFE terms (see Table 1), but exhibit a bias in opposite direction, those two methods are combined. For the direct forecast the $VAR^{int(2)}$, the AR^{SC} and the VAR^{int}_{Gets} are combined since those methods perform best in RMSFE. $VAR^{int(2)}$ and AR^{SC} have a positive bias whereas VAR^{int}_{Gets} exhibits a negative bias. For the indirect forecast combining does indeed lead to some improvement over the best respective forecast, whereas for the direct method the combined forecast is similar to the best one. However, the direct forecast still performs best overall for the forecast combinations chosen.

HICP excluding energy and unprocessed food 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 5.

Table 5: **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 1998(1),...,2000(12)**

horizon	1		6		12	
method	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}^{core}_{sub}$	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}^{core}_{sub}$	direct $\Delta_{12}\hat{p}^{core}$	indirect $\Delta_{12}\hat{p}^{core}_{sub}$
RW	0.103	0.105	0.551	0.565	1.068	1.097
$VAR^{Ph(12)}$	0.065	0.060	0.244	**0.226**	0.584	0.570
$VAR^{Int(2)}$	0.078	0.075	0.281	0.264	0.525	**0.490**
AR^{SC}	**0.061**	**0.055**	**0.237**	**0.226**	**0.520**	0.501
VAR^{Int}_{Gets}	0.065	0.111	0.265	0.449	0.554	0.842

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest RMSFE per column

Here the results show a different pattern. Three out of five methods exhibit a better accuracy for aggregating the subcomponent forecasts for a forecast horizon of one month, i.e. all methods except for the RW and the VAR^{Int}_{Gets}. A similar pattern is found for the 6 and 12 months ahead forecasts. Similar results are also exhibited by the average RMSFE (ARMSFE) over all forecast horizons in Table 6. The results for VAR^{Int}_{Gets} in Table 5 also show that for the three components the varying specification across components in terms of variables chosen does not improve the forecast accuracy of aggregating the subcomponent forecasts. Figures 6 and 7 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, the $VAR^{Ph(12)}$ and the $VAR^{int(2)}$ methods exhibit very similar forecasts of the direct and indirect method for year-on-year 'core' inflation for one month ahead forecasts (Figure 6). In contrast, for 12 months ahead forecasts the difference is up to 0.4 percentage points for these three models (Figure 7). There appears to be a similar difference in the forecasts by the direct and indirect method for the VAR^{Int}_{Gets}, in which case the direct method provides a more accurate forecast in RMSFE terms. Overall, these findings confirm that the better RMSFE accuracy of the majority of models for the indirect method of aggregating subcomponent forecasts matters in terms of the actual 'core' inflation forecast. This 'core' inflation series in-

Table 6: **Average relative forecast accuracy over all forecast horizons, average RMSFE and average MFE of year-on-year HICP inflation excluding energy and unprocessed food in percentage points, Recursive estimation samples 1992(1) to 1998(1),...,2000(12)**

criterion	ARMSFE		AMFE	
method	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$	direct $\Delta_{12}\hat{p}^{agg}$	indirect $\Delta_{12}\hat{p}^{agg}_{sub}$
RW	0.591	0.607	-0.517	0.539
$VAR^{Ph(12)}$	0.291	0.278	0.050	-0.064
$VAR^{Int(2)}$	0.294	0.275	-0.003	0.070
AR^{SC}	**0.267**	**0.262**	**0.005**	**0.060**
VAR^{Int}_{Gets}	0.293	0.471	-0.094	-0.172

Note: super and subscripts indicate model selection procedure, SC: Schwarz criterion, Ph(12): Phillips curve model including inflation and unemployment, 12 lags, Int(2): model including international variables in addition to domestic ones, 2 lags, Gets: model selection with PcGets (Hendry & Krolzig, 2001a), bold numbers: indicate lowest ARMSFE and AMFE (in absolute terms) per column

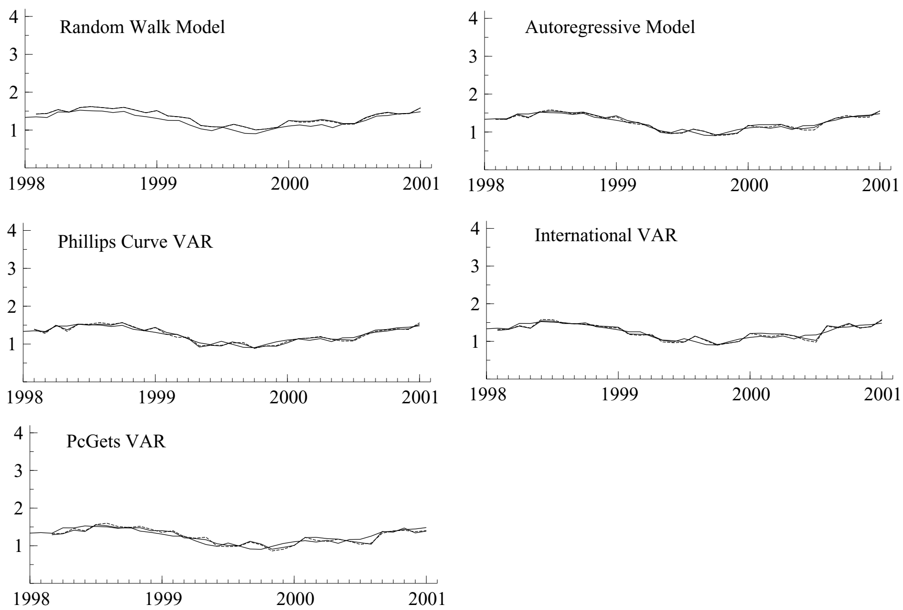


Figure 6: Year-on-year inflation rate of HICP excluding unprocessed food and energy and forecasts in %, 1 month ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

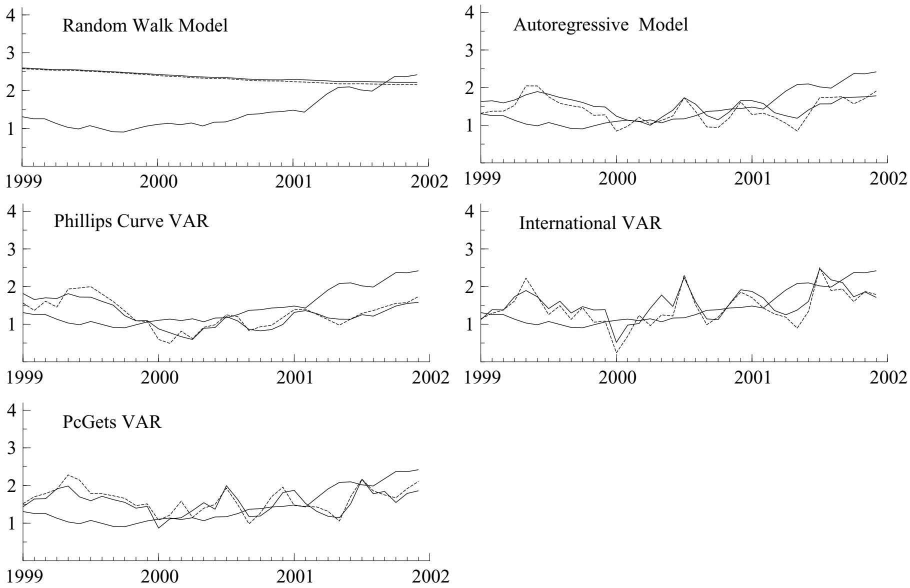


Figure 7: Year-on-year inflation rate of HICP excluding unprocessed food and energy and forecasts in %, 12 months ahead, solid: actual, dotted: aggregate forecast, dashed: aggregated subcomponent forecasts

cluding only those subindices of HICP that are less affected by shocks tends to be better forecasted by aggregating the subcomponent forecasts instead of forecasting the aggregate directly, whereas the year-on-year inflation rate of HICP total tends to be better forecasted directly at longer horizons.

6 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.

I find that it is not necessarily better to employ the indirect rather than the direct method of forecasting aggregate euro area year-on-year inflation. For many of the forecast methods considered here that are often used by practitioners and researchers, forecasting aggregate euro area year-on-year inflation directly results in higher forecast accuracy for medium-term forecast horizons of 12 months relevant for monetary policy.

The findings suggest that to forecast euro area aggregate year-on-year inflation, aggregating subcomponent forecasts has to be considered with some caution. For forecasting year-on-year inflation in the euro area the results presented raise the question whether modelling and forecasting the subcomponents is worthwhile if the forecast of the aggregate is the objective.

Although the details of the results in this study are of course specific to the empirical application of euro area inflation, the findings nevertheless point at some more general problems the forecaster may face when aggregating forecasts of disaggregate components to forecast the aggregate.

From my analysis I find the following main explanations for my results. I find that taking into account differences in the dynamic properties of subcomponents by different model specifications across subcomponents in terms of variables and/or lags do not necessarily improve the aggregate forecast. The analysis has shown that even for the VAR, where the general-to-specific model selection procedure (PcGets) selects variables to enter the respective model for each of the HICP subcomponents 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 months ahead. Furthermore, combination of different forecast methods as well as the direct and indirect forecasts is found not to improve over the best (direct) forecast 12 months ahead.

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 many shocks, e.g. the oil price shock or the shock to unprocessed food in 2000 and 2001 in the euro area, affect several or even all components of HICP over the forecast evaluation period and therefore forecast errors appear in the same direction for those components affected. Also, the extent of the effect of a shock and its dissemination across subcomponents is difficult to predict. Therefore, the forecast bias of the aggregate is not reduced, but increased by aggregating the subcomponent forecasts in this case.

Furthermore, I have investigated the forecast performance of aggregating subcomponent forecasts for another inflation measure of interest to monetary policy makers: 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. For this aggregate the majority of methods exhibits higher forecast accuracy for aggregating subcomponent forecasts. Comparing these findings with the results for overall year-on-year inflation leads to further insight into the reasons for the problems with aggregating disaggregate forecasts: Aggregating subcomponent forecasts appears to be problematic when some subcomponents are inherently difficult to forecast due to frequent shocks to the series, in case of HICP subcomponents the shocks to energy and unprocessed food prices.

Appendix A Data

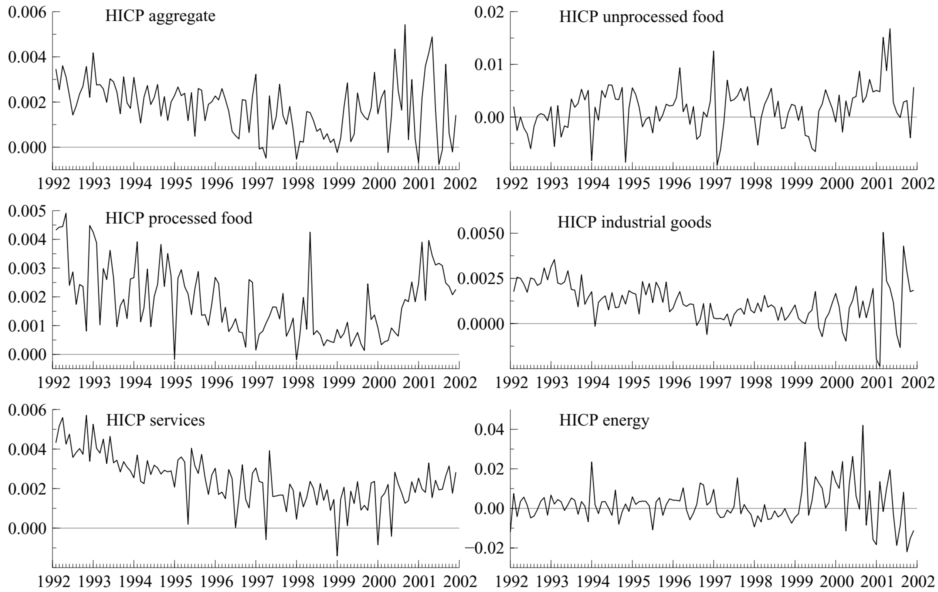


Figure 8: First differences of HICP (sub-)indices (in logarithm)

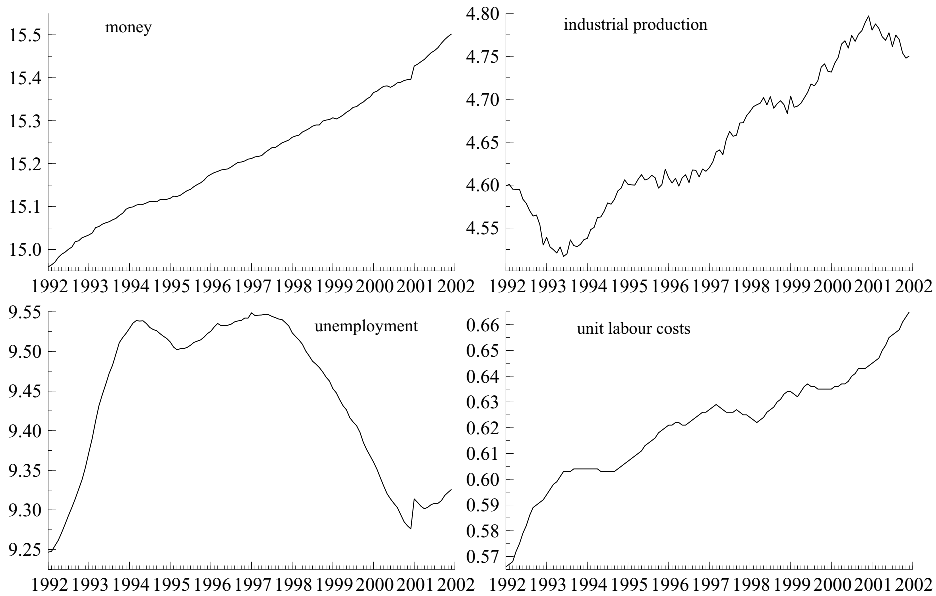


Figure 9: Money M3, industrial production, unemployment and unit labour costs (in logarithm)

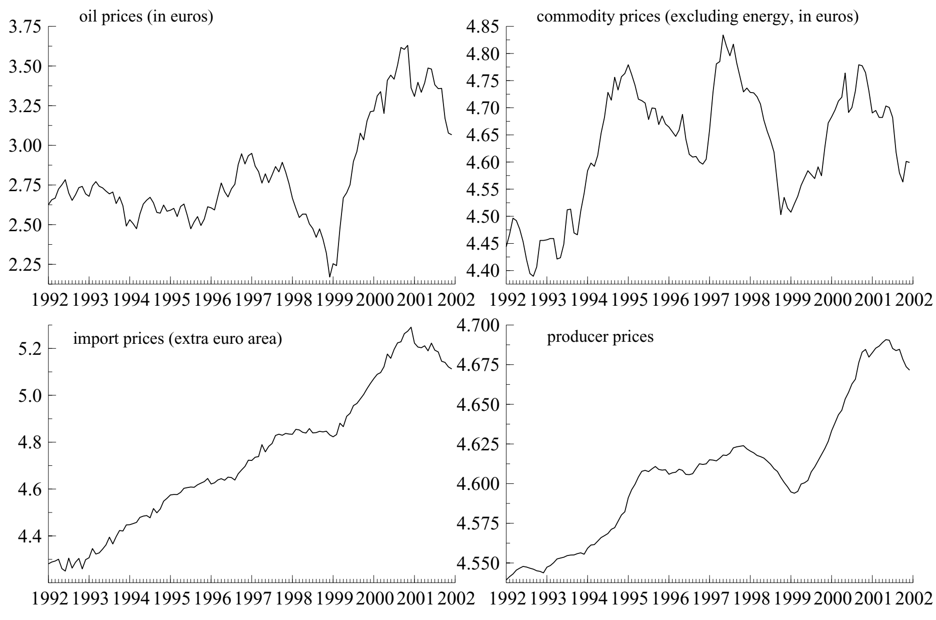


Figure 10: Oil prices (in euros), commodity prices (excluding energy, in euros), import prices (extra euro area), producer prices (in logarithm)

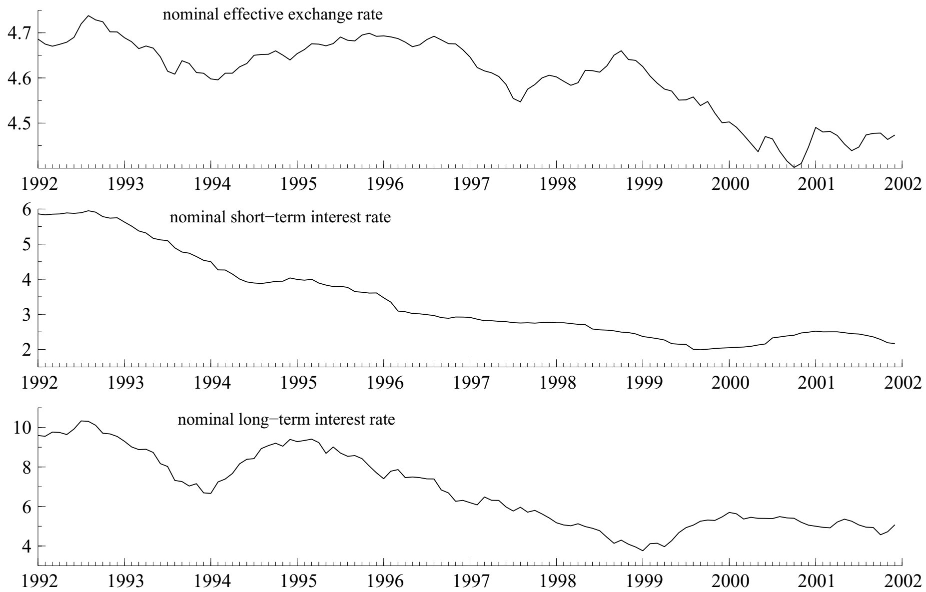


Figure 11: Nominal effective euro exchange rate, nominal short- and long-term interest rate (in logarithm except for interest rates)

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.

Clark, T. E. & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models, *Journal of Econometrics* **105**: 85–110.

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

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

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

Clements, M. P. & Hendry, D. F. (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.

Clements, M. P. & Hendry, D. F. (2002c). Pooling forecasts, *Econometrics Journal* **5**: 1–26.

Cristadoro, R., Forni, M., Reichlin, L. & Veronese, G. (2001). A core inflation index for the euro area, *Temi di Discussione 435*, Banca d'Italia.

Diebold, F. & Lopez, J. A. (1996). Forecast evaluation and combination, *in* G. Maddala & C. Rao (eds), *Handbook of Statistics, Vol. 14*, North-Holland, Amsterdam, pp. 241–268.

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

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.

European Central Bank (2000). *ECB Monthly Bulletin*, December.

European Central Bank (2003a). *ECB Monthly Bulletin*, March.

European Central Bank (2003b). ECB Press Release, The ECB's monetary policy strategy, 8 May 2003, Press Release, http://www.ecb.int.

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.

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.*

Inoue, A. & Kilian, L. (2003). On the selection of forecasting models, *ECB Working Paper 214*, European Central Bank.

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. (2002). Forecasting EMU macroeconomic variables, *Discussion Paper 3529*, CEPR.

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.

Nijman, T. & Sentana, E. (1996). Marginalization and contemporaneous aggregation in multivariate garch processes, *Jounal 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. (1999a). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, *in* R. Engle & H. White (eds), *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, Oxford University Press, Oxford, pp. 1–44.

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

Stock, J. H. & Watson, M. W. (2003). How did leading indicator forecasts do during the 2001 recession?, manuscript.

Theil, H. (1954). *Linear Aggregation of Economic Relations*, North Holland, Amsterdam.

Tiao, G. C. & Guttman, I. (1980). Forecasting contemporal aggregates of multivple 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 contemporal time series aggregates, *Communications in Statistics - Theory and Methods* (A10): 1335–1344.

West, K. D. (2001). Tests for forecast encompassing when forecasts depend on estimated regression parameters, *Journal of Business & Economic Statistics* **19**(1): 29–33.

West, K. D. & McCracken, M. W. (1998). Regression-Based tests of predictive ability, *International Economic Review* **39**(4): 817–840.

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

European Central Bank working paper series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.int).

202 "Aggregate loans to the euro area private sector" by A. Calza, M. Manrique and J. Sousa, January 2003.

203 "Myopic loss aversion, disappointment aversion and the equity premium puzzle" by D. Fielding and L. Stracca, January 2003.

204 "Asymmetric dynamics in the correlations of global equity and bond returns" by L. Cappiello, R.F. Engle and K. Sheppard, January 2003.

205 "Real exchange rate in an inter-temporal n-country-model with incomplete markets" by B. Mercereau, January 2003.

206 "Empirical estimates of reaction functions for the euro area" by D. Gerdesmeier and B. Roffia, January 2003.

207 "A comprehensive model on the euro overnight rate" by F. R. Würtz, January 2003.

208 "Do demographic changes affect risk premiums? Evidence from international data" by A. Ang and A. Maddaloni, January 2003.

209 "A framework for collateral risk control determination" by D. Cossin, Z. Huang, D. Aunon-Nerin and F. González, January 2003.

210 "Anticipated Ramsey reforms and the uniform taxation principle: the role of international financial markets" by S. Schmitt-Grohé and M. Uribe, January 2003.

211 "Self-control and savings" by P. Michel and J.P. Vidal, January 2003.

212 "Modelling the implied probability of stock market movements" by E. Glatzer and M. Scheicher, January 2003.

213 "Aggregation and euro area Phillips curves" by S. Fabiani and J. Morgan, February 2003.

214 "On the selection of forecasting models" by A. Inoue and L. Kilian, February 2003.

215 "Budget institutions and fiscal performance in Central and Eastern European countries" by H. Gleich, February 2003.

216 "The admission of accession countries to an enlarged monetary union: a tentative assessment" by M. Ca'Zorzi and R. A. De Santis, February 2003.

217 "The role of product market regulations in the process of structural change" by J. Messina, March 2003.

218 “The zero-interest-rate bound and the role of the exchange rate for monetary policy in Japan” by G. Coenen and V. Wieland, March 2003.

219 “Extra-euro area manufacturing import prices and exchange rate pass-through” by B. Anderton, March 2003.

220 “The allocation of competencies in an international union: a positive analysis” by M. Ruta, April 2003.

221 “Estimating risk premia in money market rates” by A. Durré, S. Evjen and R. Pilegaard, April 2003.

222 “Inflation dynamics and subjective expectations in the United States” by K. Adam and M. Padula, April 2003.

223 “Optimal monetary policy with imperfect common knowledge” by K. Adam, April 2003.

224 “The rise of the yen vis-à-vis the (“synthetic”) euro: is it supported by economic fundamentals?” by C. Osbat, R. Rüffer and B. Schnatz, April 2003.

225 “Productivity and the (“synthetic”) euro-dollar exchange rate” by C. Osbat, F. Vijselaar and B. Schnatz, April 2003.

226 “The central banker as a risk manager: quantifying and forecasting inflation risks” by L. Kilian and S. Manganelli, April 2003.

227 “Monetary policy in a low pass-through environment” by T. Monacelli, April 2003.

228 “Monetary policy shocks – a nonfundamental look at the data” by M. Klaeffing, May 2003.

229 “How does the ECB target inflation?” by P. Surico, May 2003.

230 “The euro area financial system: structure, integration and policy initiatives” by P. Hartmann, A. Maddaloni and S. Manganelli, May 2003.

231 “Price stability and monetary policy effectiveness when nominal interest rates are bounded at zero” by G. Coenen, A. Orphanides and V. Wieland, May 2003.

232 “Describing the Fed’s conduct with Taylor rules: is interest rate smoothing important?” by E. Castelnuovo, May 2003.

233 “The natural real rate of interest in the euro area” by N. Giammarioli and N. Valla, May 2003.

234 “Unemployment, hysteresis and transition” by M. León-Ledesma and P. McAdam, May 2003.

235 “Volatility of interest rates in the euro area: evidence from high frequency data” by N. Cassola and C. Morana, June 2003.

236 “Swiss monetary targeting 1974-1996: the role of internal policy analysis” by G. Rich, June 2003.

237 “Growth expectations, capital flows and international risk sharing” by O. Castrén, M. Miller and R. Stiegert, June 2003.

238 “The impact of monetary union on trade prices” by R. Anderton, R. E. Baldwin and D. Taglioni, June 2003.

239 “Temporary shocks and unavoidable transitions to a high-unemployment regime” by W. J. Denhaan, June 2003.

240 “Monetary policy transmission in the euro area: any changes after EMU?” by I. Angeloni and M. Ehrmann, July 2003.

241 Maintaining price stability under free-floating: a fearless way out of the corner?” by C. Detken and V. Gaspar, July 2003.

242 “Public sector efficiency: an international comparison” by A. Afonso, L. Schuknecht and V. Tanzi, July 2003.

243 “Pass-through of external shocks to euro area inflation” by E. Hahn, July 2003.

244 “How does the ECB allot liquidity in its weekly main refinancing operations? A look at the empirical evidence” by S. Ejerskov, C. Martin Moss and L. Stracca, July 2003.

245 “Money and payments: a modern perspective” by C. Holthausen and C. Monnet, July 2003.

246 “Public finances and long-term growth in Europe – evidence from a panel data analysis” by D. R. de Ávila Torrijos and R. Strauch, July 2003.

247 “Forecasting euro area inflation: does aggregating forecasts by HICP component improve forecast accuracy?” by K. Hubrich, August 2003.