Oesterreichische Nationalbank

Eurosystem

# Workshops

## Proceedings of OeNB Workshops

# *Macroeconomic Models and Forecasts for Austria*

November 11 to 12, 2004

No. 5

# Forecasting Austrian Inflation[1]

*Gabriel Moser, Fabio Rumler and Johann Scharler*

*Oesterreichische Nationalbank*

## Abstract

In this paper we apply factor models proposed by Stock and Watson (1999) and VAR and ARIMA models to generate 12-month out of sample forecasts of Austrian HICP inflation and its sub-indices processed food, unprocessed food, energy, industrial goods and services price inflation. A sequential forecast model selection procedure tailored to this specific task is applied. It turns out that factor models possess the highest predictive accuracy for several sub-indices and that predictive accuracy can be further improved by combining the information contained in factor and VAR models for some indices. With respect to forecasting HICP inflation, our analysis suggests to favor the aggregation of sub-indices forecasts. Furthermore, the sub-indices forecasts are used as a tool to give a more detailed picture of the determinants of HICP inflation from both an ex-ante and ex-post perspective.
*JEL Classification:* C52, C53, E31
*Keywords:* Inflation Forecasting, Forecast Model Selection, Aggregation

## 1. Introduction

The inflation rate is often seen as an important indicator for the performance of a central bank. Inflation forecasts are therefore an important element in the set of variables on which forward looking monetary policy decisions are based. Apart from the role of inflation forecasts as an input to monetary policy deliberations there is also an additional role for inflation forecasts in the national macroeconomic policy debate. By informing the public about likely trends in inflation the forecast can influence inflationary expectations and therefore can serve as a nominal anchor

[1] We would like to thank Sylvia Kaufmann, Ricardo Mestre, the workshop participants at the Banque de France Workshop on Inflation Forecasting and an anonymous referee for helpful comments.

for example in the wage bargaining process or for other nominally fixed contracts like housing rents, interest rates. Furthermore, since the appropriate reaction of monetary policy to inflationary pressures depends among other things on the sources of inflation, it is useful to monitor, analyze and forecast sub-indices of headline inflation that are defined according to the type of product contained in the Harmonized Index of Consumer Prices (HICP). The incorporation of information on developments in the sub-indices helps to give a more detailed picture on the sources of inflation and the propagation of shocks to inflation across product categories and time. The sub-indices covered in our analysis comprise processed food, unprocessed food, energy, non-energy industrial goods and services. In the case of the Eurosystem, a forecast of area-wide inflation is required as an input to monetary policy decisions. As area-wide inflation is an aggregation of the inflation rates prevailing in the countries of the monetary union, one way to meet this requirement is to produce inflation forecasts for the member states (for each sub index) and aggregate them to an area-wide inflation forecast.[2]

This paper compares the performance of factor models and VAR and ARIMA models for forecasting the rate of change of the Austrian HICP and its sub-indices. Furthermore, we compare the performance of HICP inflation forecasts based on "direct" modeling of the HICP with a forecast based on an aggregation of forecasts for the sub-indices.[3] The forecasts of the models with the highest predictive accuracy are then evaluated using a range of criteria that characterize *optimal* forecasts. Finally, the sub-indices forecasts with the highest predictive accuracy are used as a tool to obtain a more detailed picture of the sources of future (forecasted) inflation and past inflation forecast errors for the period 1990 to 2002.

Starting with the contribution of Stock and Watson (1999a), various authors have applied factor models to forecasting inflation. Stock and Watson (1999b) use factor models to forecast U.S. inflation. Marcellino, Stock and Watson (2000) and Angelini, Henry and Mestre (2001) evaluate the usefulness of factor models for forecasting euro area inflation. Tkacz and Gosselin (2001) evaluate factor models for forecasting inflation in Canada.

Factor models offer a convenient way to incorporate the informational content of a wide range of time series. The underlying assumption is that a small number of unobservable factors is the driving force behind the series under consideration. This is an appealing feature for forecasting purposes since it allows us to concentrate on a few common factors instead of a potentially large number of

[2] This is the approach that is currently followed in the quarterly narrow inflation projection exercises (NIPE) conducted by the Eurosystem. For a comparison of this approach with a "direct" forecast of area-wide inflation both at the level of the aggregate HICP and the sub-indices see Benalal et al. (2004).

[3] In related papers, Hubrich (2003) analyses euro area HICP sub-indices and Fritzer, Moser and Scharler (2002) consider forecasting the Austrian HICP sub-indices using time series methods.

explanatory variables. In particular, for forecasting HICP sub-indices factor models appear to be a promising tool since economic theory provides only little guidance for variable selection in this case. Hence, using factor models allows us to avoid arbitrary assumptions necessary to preserve degrees of freedom when standard time series methods are employed. On the other hand, the usefulness of other time series models, in particular VAR and ARIMA models, in forecasting inflation has been widely documented in the literature, see e.g. Hubrich (2003) and the references therein.

We find that factor models appear to possess the highest predictive accuracy for the unprocessed food, energy and industrial goods price indices. However, a check for forecastability of these variables reveals that they are close to being unforecastable which helps to explain the forecast errors made in forecasting the HICP. For processed food and the services index, the highest predictive accuracy is obtained using a combined forecast of factor and VAR models. Here the excess persistence in the forecast errors for the service price inflation forecasts stands out as a main departure from an optimal forecast. Furthermore, we find that forecasts for Austrian HICP inflation based on an aggregation of the sub-indices forecasts appears to be somewhat more accurate than the best available forecast for the HICP itself. This “indirect" approach to forecasting inflation has the additional advantage that it avoids inconsistencies between forecasts of the sub-indices inflation and headline inflation and at the same time allows a more detailed analysis of trends in inflation.

The remainder of the paper is organized as follows: Section 2 briefly discusses factor models along with the other techniques used in the forecasts and describes the forecasting procedure. Section 3 compares the forecasting performance of the models and evaluates the resulting models with the highest predictive accuracy. Section 4 concludes the paper.

# 2. Forecasting Models and Procedures

## 2.1 Forecasting Models

The goal of this paper is to evaluate forecasts for the year-on-year growth rate of the HICP index and its sub-indices. These growth rates are defined as

$$\Delta P_{i,t} = \log(P_{i,t}) - \log(P_{i,t-12}), \tag{1}$$

where $P_{i,t}$, $i = 0,...,5$, denotes the date $t$ observations of the headline HICP $i = 0$ and the sub-indices for processed food $i = 1$, unprocessed food $i = 2$, energy $i = 3$, industrial goods $i = 4$ and services $i = 5$.

The forecasting performance of the models under consideration is evaluated by comparing simulated out-of-sample forecasts. The rolling out-of-sample forecasts are carried out recursively, i.e. the models are re-estimated every period taking into account only data up to that period, as will be explained below. The out-of-sample forecast error is given by

$$u_{i,t+12} = \Delta P_{i,t+12} - \widehat{\Delta P}_{i,t+12}, \tag{2}$$

where $\widehat{\Delta P}_{i,t+12}$ is the predicted value for the year-on-year increase of index $i$.

In the case of the factor model forecasts are generated for each inflation rate as a linear projection of the change of the log price index over the next 12 months on a set of predictor variables:

$$\Delta P_{i,t+12} = \alpha_i + \sum_{h=0}^{n} \beta_{i,h} \Delta P_{i,t-h} + \sum_{l=1}^{m} \sum_{h=0}^{k} \gamma_{i,l,h} f_{l,t-h} + \varepsilon_{i,t+12}. \tag{3}$$

The change in each index over the next 12 months is explained by $n$ of its own lags plus at most $k$ lags of $m$ common factors denoted by $f_t$, $\varepsilon_{t+12}$ is an $i.i.d.$ disturbance term. In order to generate forecasts from equation (3), the factors have to be estimated. Stock and Watson show that $f_t$ can be consistently estimated by the method of principal components. Concerning the choice of the number of factors, we apply the selection criteria of Bai and Ng who specify that the number of factors, $m$, is set equal to the mode of the optimal number of factors over the estimation sample.

The second class of models considered are VAR models. In the selection of the specific VARs used in our analysis we take mainly a statistical approach. The models are selected according to pure statistical criteria instead of being derived from any theory of inflation determination. The reason is – besides the fact that the focus of the paper is not to test different models of inflation determination – a rather practical one, namely that theoretical models do not really exist for the inflation processes of the HICP sub-indices. In particular, specifying the VARs requires two decisions. First, the variables entering the VAR have to be selected. Second, the appropriate lag specification of the model has to be determined. The variables entering the six VARs for the sub-indices and the HICP are selected according to a procedure which investigates the leading indicator properties of all 179 time series in our database for the HICP sub index under consideration.[4] This

[4] The leading indicator property is assessed by the explanatory power of any of the series for the respective HICP sub index in a large number of bivariate regressions (for 1, 3, 6,

procedure is only a first step in selecting the variables because usually a larger number of variables than what is feasible to include in a VAR qualify according to our procedure. This implies that some judgment to reduce the variables to a feasible number is required which prohibits an automatic reformulation of the VAR in every period of the rolling estimations[5]. For this reason and also for the fact that using VARs with changing variables from period to period would render our forecasts rather unstable, we decided to keep the formulation of the VARs in terms of variables constant over all periods,[6] whereas the lag specification is re-optimized every period. The variables included in the VARs for the five sub-indices and the HICP aggregate which have been selected according to the procedure just described are listed in table 1[7].

As a third model class, we use ARIMA models. The specification of the ARIMA models for the five sub-indices and headline inflation, i.e. the selection of AR and MA terms as well as seasonal AR and MA terms, is also re-optimized in every period of our rolling estimation procedure. All ARIMAs are estimated in first-difference form implying that no unit root specification for the ARIMAs is required, as all indices are difference-stationary.

9, and 12 months ahead). This procedure is described in more detail in Fritzer, Moser and Scharler (2002).

[5] For example, the judgment comes into play when a few variables that are equally correlated with the sub index under consideration and which are strongly correlated among each other, one of them is selected by judgment to enter the VAR.

[6] This, however, in a strict sense violates the principle of full recursiveness of our forecasts as information of the whole sample is used in the formulation of the VARs also for the earlier periods.

[7] All VARs are estimated in first differences and all variables are in logs except the interest rate series.

*Table 1: Variables in the VAR Models (in Addition to the Respective Indices)*

| $\Delta P_0$ | $\Delta P_1$ |
|---|---|
| OeNB/ECB base rate,<br>bank deposits up to 2 years,<br>M3 | OeNB/ECB base rate,<br>bank deposits up to 2 years,<br>negotiated wages in agro-forestry,<br>price index of foreign tourist demand,<br>wholesale price index for food and beverage |
| $\Delta P_2$ | $\Delta P_3$ |
| total orders in manufacturing,<br>unemployment in building and construction, wholesale price index for feed barley,<br>negotiated wages in agro-forestry | OeNB/ECB base rate,<br>industrial production,<br>bank deposits up to 2 years,<br>M3,<br>exports of intermediate goods |
| $\Delta P_4$ | $\Delta P_5$ |
| M3,<br>demand deposits,<br>producer price index for pulp wood,<br>exports of final goods | total credit,<br>bank deposits up to 2 years,<br>M3,<br>wholesale price index for mineral oil |

## 2.2 Data and Forecast Procedure

Our data set consists of 179 macroeconomic and financial time series of monthly frequency, beginning in 1980:1 and ending in 2002:12. This yields a total of 276 observations for each series. The data are seasonally adjusted and outliers are removed. For the estimation the series are differenced in order to induce stationarity. Since the HICP and its sub-indices are only available from 1987:1 on, we extrapolate the series backwards until 1980:1 in order to increase the number of observations (see Appendix A). Furthermore, we remove breaks from the processed food and the industrial goods index together with the corresponding breaks in the HICP before we forecast those series (see Appendix B). The HICP and its sub-indices are also seasonally adjusted.

For the evaluation and comparison of the different model classes we construct a series of 12-step-ahead out-of-sample forecasts where the models are estimated with a rolling split of the estimation and forecasting periods. In moving forward the rolling procedure the models are not only re-estimated each period but also their lag specifications are re-optimized after each step. This is done by estimating a large number of possible lag combinations for each VAR, ARIMA and factor model out of which the model with the best fit according to the Schwarz information criterion is selected.[8] The selected model is then used to produce a 12-step-ahead forecast where only the last forecast value, i.e. the 12th-step-ahead forecast, is used for the forecast evaluation. In the next step the estimation sample is moved one period forward, again a large number of different lag specifications are estimated, the optimal model is selected and used to produce another 12-step-ahead forecast where again only the last value is stored for the forecast evaluation.[9] The procedure continues until the last 12-step-ahead forecast has reached the end of the sample range.

Specifically, we start with the estimation period 1980:1 to 1989:1 and forecast the values for the period 1989:2 to 1990:1. The forecast for 1990:1 is the first to be used for the evaluation.[10] Next, the models are estimated for the period 1980:1 to 1989:2 to produce a forecast for 1989:3 to 1990:2 where only the last value is stored for later evaluation, and so on. By stacking all the stored values we obtain a series of 12-step-ahead forecasts for HICP inflation and its sub-indices ranging from 1990:1 to 2002:12 – each derived from a different forecast – which are then compared with the true values and the forecasts of the other models.

# 3. Forecasting Model Selection and Evaluation

This section has three goals: first, the forecasts of factor models and VAR and ARIMA models of the HICP and its sub-indices are compared. Based on this

[8] Concerning the VARs, a total of 451 specifications are estimated each period which include (not all possible but most relevant) lag combinations from a minimum of 4 up to a maximum of 14 lags. In the case of the ARIMAs a total number of 676 specifications are estimated each period including all possible combinations of AR and MA terms up to 12 lags as well as seasonal AR and MA terms at the 12$^{th}$ lag.

[9] The fact that the specifications are re-optimized after each period also implies that two consecutive forecasts may be based on different models, which has the potential to make the series of forecasts more variable. However, in our estimations – except for only a few periods – this did not turn out to be a major problem.

[10] We chose the minimum estimation sample to range from 1980:1 to 1989:1 because, given the large number of coefficients to be estimated for some specifications, a fairly large number of observations is required to deliver reliable estimates. Furthermore, as noted by Ashley (2003), a sufficient number of observations in the validation period, preferably above 100, is necessary to establish significant differences in predictive accuracy.

comparison possibilities for forecast combination are considered. Second, a distinct approach to generate forecasts of the HICP is examined namely a “indirect" forecast based on an aggregation of forecasts of the HICP sub-indices. Both steps are conducted with the goal of arriving at a specification for forecasting Austrian inflation that is characterized by highest predictive accuracy. Third, this specification to forecast the HICP and its sub-indices is evaluated and the minimized forecast errors are used for an ex-ante and ex-post assessment of Austrian inflation during the period 1990–2002.

## 3.1 Forecast Comparison and Forecast Combination

The comparisons are based on a common descriptive statistic for predictive accuracy, the mean squared error (MSE), a test for differences in predictive accuracy and a test for forecast encompassing. As the latter testing principle is related to the concept of forecast combination we also compute the MSE for combined forecasts where appropriate. Factor models are used as benchmark for comparing predictive accuracy. This choice appears inconsequential, i.e. does not appear to prevent efficient forecasting model selection, as it only entails that we do not compare the VAR and the ARIMA model with respect to their relative predictive accuracy.

As a descriptive statistic for the gain of using factor models we compute

$$Gain_{1,j} = 100 * \left(\frac{MSE_j - MSE_1}{MSE_j}\right) \tag{4}$$

where $MSE_1$ is the mean squared error of the factor model forecast and $MSE_j$ the competing models forecast ( $j = 2$ denotes the VAR model and $j = 3$ denotes the ARIMA model). As a “rule of thumb" a model is considered to possess higher predictive accuracy if the gain is above 10%, a choice that can be found in the literature on forecast comparisons (see e.g. Marcellino et al., 2000).

Furthermore, as formal statistical testing for relative predictive accuracy is usually recommended (see Fildes and Stekler, 2002) we make use of the test statistic of Harvey, Leybourne and Newbold (1997) which is a modified version of the widely used statistic of Diebold and Mariano (1995). This statistic is applied to test the null hypothesis of equal predictive accuracy between the factor model forecast as the benchmark model and the VAR and the ARIMA model forecast. The modification proposed by Harvey, Leybourne and Newbold should reduce somewhat the size distortion of the Diebold-Mariano test that is present when long horizon multi-step forecasts are compared.

The distribution of this statistic is an issue of debate. Harvey, Leybourne and Newbold suggest to use the Student distribution with $N-1$ degrees of freedom. Clark and McCracken (2002) show for the case of multi-step forecasts that this is no longer appropriate when forecasts are derived from nested models. We follow the suggestion of Harvey, Leybourne and Newbold and compare the values of this statistic with the critical values of a Student distribution with 155 degrees of freedom associated with a 10% and 5% confidence level. The critical values are 1.66 and 1.98.

The decision rule based on the descriptive and the test statistic should ideally lead to one model with higher predictive accuracy than all competing models for each index. The next step consists of determining whether these models also encompass their competitors. Forecast encompassing is given when a forecast already incorporates all the relevant information of a competing forecast. The concept of forecast encompassing is related to the idea of forecast combination. If a forecast does not encompass the competing forecast then there might exist a linear combination of the two forecasts with further improved predictive accuracy.

We make use of the encompassing test statistic of Harvey, Leybourne and Newbold (1997). With respect to the distribution of this statistic the same issue arises as in the case of tests for equal predictive accuracy (see Clark and McCracken, 2002). We again follwow Harvey, Leybourne and Newbold and compare the values of this statistic with the critical values of a Student distribution with 155 degrees of freedom associated with a 10% and 5% confidence level. As this is a one-sided test the critical values are 1.29 and 1.66.[11]

Based on the results of the encompassing tests we then employ the variance-covariance approach to forecast combination proposed by Bates and Granger [4]. This approach applies the logic of portfolio optimization to forecast combination. Consider the following linear combination of the forecast of the model with higher predictive accuracy $\widehat{\Delta P}_{i,s,t+12}$ and a competing forecast $\widehat{\Delta P}_{i,j,t+12}$ for the inflation rate of index $i$:

$$\widehat{\Delta P}^{c}_{i,s,j,t+12} = \omega \widehat{\Delta P}_{i,s,t+12} + (1-\omega)\widehat{\Delta P}_{i,j,t+12}. \tag{5}$$

Given that both the forecast of the model with higher predictive accuracy and the competing forecast are unbiased one can show that the weight $\omega^*$ which minimizes the forecast error variance of the combined forecast $\widehat{\Delta P}^{c}_{i,s,j,t+12}$ is given by

[11] Details on the definition and application of the tests for comparing predictive accuracy and forecast encompassing can be found in Appendix C.

$$\omega_i^* = \frac{var(u_{i,j,t+12}) - cov(u_{i,s,t+12}, u_{i,j,t+12})}{var(u_{i,j,t+12}) + var(u_{i,s,t+12}) - 2cov(u_{i,s,t+12}, u_{i,j,t+12})}, \tag{6}$$

where $u_{i,j,t+12}$ and $u_{i,s,t+12}$ are the forecast errors of the two model. The mean squared error of the combined forecast associated with the optimal combining weight $\omega_i^*$ is denoted as $MSE_{i,s,j}^C$. This measure has the property that $MSE_{i,s,j}^C \leq min(MSE_{i,s}, MSE_{i,j})$.

### *3.1.1 Comparing Factor Models with VAR and ARIMA Models*

We begin with computing the MSEs for all indices and for each forecasting model and the corresponding gains in using factor models. After verifying the stationarity of the loss differential sequences the null hypothesis of equal predictive accuracy of the factor model compared to the VAR and ARIMA models and the null hypothesis that the resulting models with higher predictive accuracy encompass the competing forecasts are tested. Finally, if encompassing can be rejected, optimal combining weights, the corresponding combined forecasts, its MSE and the associated gain in predictive accuracy of a combined forecast compared to the forecast of the model with higher predictive accuracy are computed. The results are shown in table 2.

One immediate result is that ARIMA models do not appear to possess higher predictive accuracy for any of the indices. Furthermore, encompassing of the ARIMA model by the factor model or the VAR model cannot be rejected for any of the indices. Therefore ARIMA models do not appear to perform well relative to the two other models, leaving the choice of using factor model forecasts, VAR model forecasts or combined forecasts of these two models.

The factor model for the HICP inflation rate $\Delta P_0$ seems to work somewhat better than the VAR model with a gain of 19%. This gain is not significant, however. Encompassing of the VAR model forecast by the factor model forecast cannot be rejected.

*Table 2: Forecast Performance of the Factor, VAR and ARIMA Models*

| | $\Delta P_0$ | $\Delta P_1$ | $\Delta P_2$ | $\Delta P_3$ | $\Delta P_4$ | $\Delta P_5$ |
|---|---|---|---|---|---|---|
| $MSE_1$ | 0.35 | 0.76 | 9.67 | 21.2 | 0.35 | 0.42 |
| $MSE_2$ | 0.43 | 0.70 | 11.4 | 26.7 | 0.56 | 0.48 |
| $MSE_3$ | 0.80 | 1.55 | 11.3 | 38.2 | 0.49 | 0.80 |
| $Gain_{1,2}$ | 19 | –8 | 16 | 21 | 38 | 12 |
| $Gain_{1,3}$ | 56 | 51 | 15 | 45 | 29 | 47 |
| $DM_{1,2}^{mod}$ | 1.29 | –0.3 | 2.07 ** | 2.00 ** | 2.34 ** | 1.14 |
| $DM_{1,3}^{mod}$ | 1.92 * | 1.74 * | 1.28 | 3.07 ** | 2.07 ** | 2.86 ** |
| $HLN_{s,j}$ | 0.69 | 1.95 ** | 0.18 | 0.33 | –1.84 | 2.07 ** |
| $HLN_{s,3}$ | –1.94 | –2.10 | 0.79 | –0.51 | –1.28 | –2.11 |
| $MSE_{s,j}^{C}$ | - - | 0.62 | - - | - - | - - | 0.37 |
| $Gain_{s,j}^{C}$ | - - | 12 | - - | - - | - - | 13 |

*Notes:* $DM_{1,j}^{mod}$ *denotes the modified Diebold-Mariano test statistic.* $HLN_{s,j}$ *denotes the encompassing statistic of Harvey, Leybourne and Newbold.* ** *and* * *indicate rejection of the null at the 5% and 10% level.*
*Source: Authors' calculations.*

For the processed food price inflation rate, $\Delta P_1$, the VAR model produces the smallest MSE of all models. However, the gain compared to the factor model is only 8% and not significant. This is an example for two models being within a "demilitarized zone" (see Kunst, 2003) within it is not possible to discriminate between models due to their high degree of similarity in performance. Nevertheless, we test whether the VAR model does encompass the factor model. As encompassing can be rejected we compute a combined forecast that appears to improve predictive accuracy compared to the VAR model, with a gain of 12%. The factor model for the unprocessed food price inflation rate, $\Delta P_2$, outperforms the VAR model significantly with a gain of 16%. Encompassing cannot be rejected. The inflation rates of energy prices, $\Delta P_3$, and industrial goods prices, $\Delta P_4$, are also forecast best by the factor model with large and significant gains compared to the VAR models. Encompassing cannot be rejected. Finally, the factor model forecasts of the inflation rate of services prices, $\Delta P_5$, seem to outperform the VAR forecast. However, the difference in the MSE is not significant. Encompassing can be rejected for the VAR model and the corresponding combined forecast of the factor

model forecast and the forecast of the VAR appears to improve predictive accuracy with a gain of 13%.

Overall, the factor model appears to produce highest predictive accuracy in terms of a lower MSE for forecasting unprocessed food, energy and industrial goods. Encompassing can be rejected for the processed food price and the services price index. The combination of the two models forecasts for these indices seem to produce forecasts with further improved predictive accuracy.

Therefore, the specification for forecasting the sub-indices of the Austrian HICP with highest predictive accuracy consists of factor models for forecasting unprocessed food price inflation, energy price inflation and industrial goods price inflation and a combined forecast of the factor model and the VAR model for forecasting processed food and service price inflation. With respect to forecasting the HICP, the factor model displays higher predictive accuracy. However, there exists another approach to forecast the HICP that can potentially produce forecasts with still higher predictive accuracy. This approach consists of a contemporaneous aggregation of forecasts of the sub-indices to a forecast of the HICP, an issue that will be addressed in the next subsection.

### *3.1.2 Comparing the Direct and the Indirect Approach to Forecasting the HICP*

The fact that the HICP is a weighted average of its sub-indices opens up another possibility to arrive at forecasts for the HICP, namely the contemporaneous aggregation of the forecasts of the sub-indices to a forecast of the HICP. Following the terminology in Hubrich (2003) this approach is referred to as the indirect approach while forecasting the HICP itself is considered the direct approach. Theoretically, if the data generating processes of the sub-indices are known, the indirect approach should yield a lower MSE since it is based on a larger information set. However, if the data generating process is not known, as is the case in this study, there are no reasons based on statistical theory to favor either approach (for a survey on the theoretical aspects of forecast aggregation see the paper of Hubrich). The following equation relates the HICP to its sub-indices:[12]

$$P_{0,t} = \sum_{i=1}^{5} w_{i,t} P_{i,t}. \tag{7}$$

[12] Due to the method of aggregation there may be small deviations between the weighted sum of the sub-indices and the HICP as provided by the statistical office of Austria. During the period 1990–2002 the average discrepancy is 0.03 percentage points.

The weights $w_{i,t}$ add up to unity and represent expenditure shares of the representative consumers consumption basket as measured by the statistical office of Austria. Under the assumption of constant weights this translates into

$$\hat{u}^{agg}_{0,t+12} = \sum_{i=1}^{5} w_i \hat{u}^{*}_{i,t+12}. \tag{8}$$

The forecast error of the indirect approach $\hat{u}^{agg}_{0,t+12}$ is equal to the weighted sum of the forecast errors $\hat{u}^{*}_{i,t+12}$ of the sub-indices forecasts with highest predictive accuracy determined in the previous section.

In our application the weights are time varying which implies that the weighted sum of the first differences of the components of the HICP is not necessarily equal to the first difference of the HICP. As the future weights are in general not known, we use a random walk forecast for the weights twelve months ahead. Since year-on-year changes in the weights are usually small, this method does not affect the forecast and therefore the comparison in a major way, indicating that time variation in the weights is not an important factor for indirect HICP forecasting at the 12-month horizon. Given the forecast of the future values of the weights we generate the HICP forecasts based on the indirect approach and calculate the associated forecast errors. This forecast is compared to the direct HICP forecasting model with higher predictive accuracy determined in the previous section (the factor model) by computing their MSEs, the gain and the test statistics for comparing predictive accuracy and encompassing. Note however, that in this case it is unclear whether there exists a nesting relationship between the direct and the indirect approach. Nevertheless, we report the $DM^{mod}$ and $HNL$ statistics and compare them with the same critical values as in the previous section. table 3 shows the results.

*Table 3: Forecast Performance of the Direct and the Indirect Approach*

| | $\Delta P_0$ |
|---|---|
| $\text{MSE}_1$ | 0.35 |
| $\text{MSE}_{agg}$ | 0.31 |
| $\text{Gain}_{1,agg}$ | –11 |
| $\text{DM}^{mod}_{1,agg}$ | –0.98 |
| $\text{HNL}_{1,agg}$ | –0.29 |

*Notes:* $DM^{mod}_{1,j}$ *denotes the modified Diebold-Mariano test statistic.* $HLN_{s,j}$ *denotes the encompassing statistic of Harvey, Leybourne and Newbold.* $^{**}$ *and* $^{*}$ *indicate rejection of the null at the 5% and 10% level.*
*Source: Authors' calculations.*

Based on the descriptive statistic it appears that the indirect approach to forecasting the Austrian HICP produces forecasts with higher predictive accuracy, with a gain of 11%. However, according to the modified Diebold-Mariano test the difference of the MSEs is not significant. Furthermore, the null hypothesis that the indirect approach encompasses the direct approach cannot be rejected which renders forecast combination between the two approaches irrelevant. Therefore, a decision rule based on statistical criteria would tend to select the indirect approach as it produces highest predictive accuracy.

A further non-statistical reason to select this approach to forecasting HICP inflation is that the direct and the indirect approaches will usually give different forecasts at any given point in time. If the direct approach is used it is likely that the forecast of the HICP and the forecasts of the sub-indices are inconsistent. This natural disadvantage of the direct approach would have to be compensated by visible gains to predictive accuracy, something which does not appear to be the case for the sample under study. However, the corresponding natural advantage of the indirect approach is based on the presumption that forecasts of the sub-indices of the HICP have an intrinsic value beyond being instrumental for forecasting the HICP.

This intrinsic value consists of a more detailed picture of expected trends in inflation which can be useful for a forward looking monetary policy. Such an ex-ante assessment of HICP inflation requires sub-indices forecasts with a high degree of predictive accuracy. Furthermore, the disaggregated approach can also help to identify the sources of past shocks to HICP inflation. An example for the use of sub-indices for an ex-post evaluation of shocks to Euro Area HICP inflation since the beginning of stage III of EMU is given in ECB (2002), p. 34.

## 3.2 Evaluating the Models with Highest Predictive Accuracy

In the previous two sections a specification for forecasting the sub-indices and the HICP has been determined that is characterized by highest predictive accuracy. It consists of factor models for unprocessed food, energy and industrial goods price inflation and combined forecasts of factor and VAR models for processed food and services price inflation. The preferred forecast for the HICP is obtained using the indirect approach. The forecasts along with the actual inflation rates are shown in charts 17–22.

The next step is to check whether the resulting forecasts with highest predictive accuracy satisfy a range of criteria which characterize *optimal* forecasts, as listed in Diebold and Lopez (1996). If departures from optimality are detected, it may be possible to improve these forecasts accordingly. The first two evaluation criteria are whether the forecasts are efficient and unbiased. This can be checked by running the regression $\hat{u}^*_{t+12} = \beta_1 + \beta_2 \Delta \hat{P}^*_{t+12} + \varepsilon_{t+12}$, where $\hat{u}^*_{t+12}$ is the forecast

error of the model with the highest predictive accuracy, $\Delta\hat{P}^{*}_{t+12}$ the corresponding forecast and $\varepsilon_{t+12}$ an $i.i.d.$ error term. If $\beta_1$ is insignificant this indicates unbiasedness, while an insignificant $\beta_2$ coefficient indicates efficiency as the forecast error is unrelated to the forecast itself. Furthermore, optimal k-step forecasts errors should display at most $(k-1)$-dependence. For our application this implies that there should be no significant autocorrelation at any lag greater than lag 11. This can be checked by examining the autocorrelation function of the forecast error series and comparing the autocorrelations with the confidence bound $+/-2/\sqrt{N}$. It is also of interest to test for normality of the distribution of the forecast errors which can be done with the Jarque-Bera test.[13]

Another evaluation criterion of interest is whether the inflation rates of the sub-indices and the HICP are actually forecastable conditional on our dataset and our models with highest predictive accuracy. Determining the degree of forecastability in particular of the sub-indices is useful as it helps to explain the errors in forecasting HICP inflation. A common measure of forecastability which is mentioned by Diebold and Lopez (1996) is the statistic $G = 1-(var(\hat{u}_t)/var(y_t))$ where $\hat{u}_t$ is the forecast error and $y_t$ is the actual value of the series to be forecast. This statistic has the form of the $R^2$ of a linear regression, i.e. it indicates the proportion of the variance explained by the model of the total variance of the series. A low value indicates a low degree of forecastability. The results of the checks for unbiasedness, efficiency, departure from $(k-1)$-dependence, normality and forecastability are shown in table 4.

[13] This test does not account for the serial correlation present in the forecast error series.

*Table 4: Criteria for an Optimal Forecast*

| | $\Delta P_0$ | $\Delta P_1$ | $\Delta P_2$ | $\Delta P_3$ | $\Delta P_4$ | $\Delta P_5$ |
|---|---|---|---|---|---|---|
| Unbiasedness | √ | √ | √ | | | |
| Efficiency | √ | √ | √ | | √ | |
| (k-1) independence | | √ | √ | | √ | |
| Normality | √ | √ | √ | √ | | √ |
| Forecastability | 0.31 | 0.39 | 0.02 | –0.01 | 0.04 | 0.67 |

*Note: √ indicates fulfillment of the criteria.*
*Source: Authors' calculations.*

The estimates for $\beta_1$ and $\beta_2$ indicate that the forecasts for energy, industrial goods and services price inflation rates are biased. As the sample means of the corresponding forecast errors are 0.8, 0.2 and 0.1 percentage points, it appears that with the exception of the energy index the biases are minor. The test for efficiency indicates that the forecasts of the services index and of the energy index are not efficient. The inspection of the autocorrelation functions of the forecast error series show that the acf of the processed food, energy and services indices indicate substantial serial correlation beyond lag 11. The other indices' acf dies out smoothly until lag 11 with no substantial autocorrelation thereafter (see charts 5 to 10). Since the services index has the largest weight in the HICP, it could be useful to try to exploit this regularity in the forecast error for improving predictive accuracy. The tests for normality of the distribution of the forecast errors indicate that with the exception of the industrial goods price inflation forecast error all forecast error distributions are normal at the 5% confidence level (see charts 11 to 16 for the forecast error distributions and their moments).

The check for forecastability shows that the indices for unprocessed food and energy prices are essentially unforecastable with the methods and data employed in this paper. Industrial goods price inflation also appears very difficult to forecast. Forecasting processed food prices is considerably more successful. The highest degree of forecastability is found for the services price index with a value of 0.67. As the forecast of the HICP is produced using the indirect approach, the medium degree of forecastability of the HICP inflation rate also reflects the different degrees of forecastability of the sub-indices.

A further important method to assess the quality of a forecast is a visual inspection against the actual series and a visual inspection of the forecast error. Since the forecasts display considerable short run variation, a centered three month moving average of the actual series, the forecasts and the forecast errors is used in

order to facilitate the identification of patterns (see charts 17 to 22 and 23 to 28). In the interest of brevity, only the HICP forecast inspection is described here.

The HICP inflation forecast underpredicts inflation considerably from mid 1990 to 1993, followed by a period of very good forecasting performance from 1994 to 1997. The following period is characterized by a considerable overprediction of inflation in 1999 followed by an underprediction in 2000 and 2001. The visual inspection of the actual series suggests major shifts in the trend of inflation in 1993, 1999 and 2001. The model appears to predict those turning points in inflation well, albeit with a lag. The graph of the HICP forecast error shows the same information in a different representation. Note that the forecast error series crosses the zero line often, which is considered a desirable property of a forecast error.

It was mentioned that forecasts of the sub-indices of the HICP can help to identify the sources of HICP inflation both from an ex-ante and ex-post perspective. As the forecast of the HICP is based on an aggregation of sub-indices forecasts, a decomposition of the forecast as well as a decomposition of the forecast errors can be obtained using equations (7) and (8).[14] A visual representation of the decomposition of the HICP forecast and the HICP forecast errors at annual frequency is given in charts 29 and 30.

For the HICP forecast decomposition the following picture emerges: Until 1994 the models predicted a rather stable HICP inflation at close to 2% with the exception of 1992 where forecasted inflation was lower. Throughout, forecasted inflation was mainly driven by increases in services prices, a feature that is also maintained in the years after 1994, where forecasted HICP inflation receded to around 1.5%. The stability in the contribution of service sector inflation to forecasted headline inflation can be attributed to offsetting tendencies of a trend decline in forecasted services inflation and a trend increase in the weight of services in the consumer basket (from 36% in 1990 to 45% in 2002). The other indices did not contribute much to forecasted inflation, either due to their small weight and/or due to a small forecasted inflation. The year-to-year variation in forecasted inflation can be mainly attributed to the contribution of the forecast of energy price inflation.

Turning to the HICP forecast error decomposition, the higher than expected inflation during the period from 1990 to 1993 was broadly based across goods categories, with large contributions of unprocessed food in 1990–91 and of services and industrial goods throughout. The unexpectedly low inflation between 1997 and 1999 was related to unexpectedly low industrial goods price inflation and unexpectedly low energy price inflation. The unexpectedly high inflation in 2000 emanated almost exclusively from the energy category, while in the following year inflation rates in almost all categories were underestimated. In the years 1995 to

[14] Note that this exercise is stylized in the sense that the adjustments to the processed food and industrial goods indices described in appendix B are not taken into account.

1996 a considerable underprediction of energy prices did offset overpredictions of inflation in other components. Recalling the result that energy price inflation is essentially unforecastable this indicates that a low degree of forecastability is not a sufficient condition for dismissing the attempt to forecast a variable.

To sum up, the ex-ante analysis of expected trends in Austrian inflation revealed that based on the models selected, a forward-looking decision-maker would have attributed most of inflation to increases in services prices, and she would have predicted a significant shift in the level of inflation in 1994 and 1995, partly explained by lower forecast energy and industrial goods price inflation. The errors implied by that ex-ante assessment were widely spread across goods categories at the beginning of the nineties while a strongly oscillating oil price was the dominating cause of over- and underpredictions of Austrian HICP inflation at the turn of the century.

Note that this is a stylized analysis designed to give an example for the use of HICP sub-indices for obtaining a more detailed picture of trends in inflation and not a description of the information available to decision makers in the past. The reason is that the sequential forecast model selection procedure applied above (and the variable selection procedure for the VARs) uses information from the whole period from 1990 to 2002. A more realistic exercise would require recursive forecasting together with recursive forecasting model selection. This is beyond the scope of this paper.

# 4. Conclusions

In this paper we take a comprehensive approach to forecast Austrian inflation at the 12-month horizon by forecasting aggregate HICP as well as 5 sub-indices. The simulated recursive out-of-sample forecasting exercise together with the forecasting model selection procedure suggest that factor models are useful for forecasting the sub-indices of the HICP. In two cases, predictive accuracy can be further improved by combining factor models with VAR models. An aggregation of sub-indices forecasts yields a somewhat higher predictive accuracy than a forecast of the HICP, with the additional advantage of consistency between the forecast of the HICP and the forecasts of the sub-indices. Furthermore, those forecasts can be used to give a more detailed picture of the determinants of HICP inflation both from an ex-ante and ex-post perspective. The analysis of the degree of optimality of the forecasts with highest predictive accuracy reveals some departures from optimality along several dimensions. The analysis of the forecastablity of the indices suggests that the specification with highest predictive accuracy obtainable from the models considered is still not able to forecast energy prices and unprocessed food prices. Industrial goods price inflation also appears difficult to forecast. However, in the case of energy price inflation the forecast

errors tended to reduce the error of the HICP forecast by offsetting errors in other sub-indices.

The recursive out-of sample forecasting procedure is designed to simulate the problem of a forecaster of Austrian inflation in real time. However, this situation is far more complex than can be replicated in such an exercise. Conducting a real time forecast usually entails, besides selecting the optimal model, the use of personal judgment of the forecaster which is based on her expertise and experience. Apart from that, it is not guaranteed that the models, even with the highest predictive accuracy, use the information available in the data in an optimal way. Hence, the difference between our exercise and the job of forecasting inflation in real time is that in the latter case, the numeric results of the models, the judgment of the forecaster and additional information on future likely events affecting inflation, such as planned fiscal measures by the government or likely developments of raw material prices, interest rates and exchange rates derived from financial market prices, all combine to produce a more accurate forecast. This implies that predictive accuracy which has been the focus of this paper, although being vital, is not the only determinant for selecting the type of models to be used in forecasting Austrian inflation.

## References

Angelini, E., Henry, J. and Mestre, R. (2001), "Diffusion index-based inflation forecasts for the euro area", Working Paper 61, European Central Bank.

Ashley R. (2003), "Statistically significant forecasting improvements: how much out-of-sample data is likely necessary?", International Journal of Forecasting 13, pp. 229–240.

Bai, J. and Ng, S. (2002), "Determining the number of factors in approximate factor models", Econometrica 70, pp. 191–221.

Bates, J. M. and Granger, C. W. (1969), "The combination of forecasts", Operations Research Quarterly 20, pp. 451–468.

Benalal, N., Diaz del Hoyo, J.L., Landau, B., Roma, M. and F. Skudelny (2004) "To aggregate or not to aggregate? Euro area inflation forecasting", ECB Working Paper 374.

Clark, T. E. and McCracken, M. W. (2002), "Evaluating long-horizon forecasts", Research Working Paper 01–14, Federal Reserve Bank of Kansas City.

Diebold, F. X. and Mariano R. S. (1995), "Comparing predictive accuracy", Journal of Business and Economic Statistics 13, pp. 253–263.

Diebold, F. and Lopez, J. (1996), "Forecast evaluation and combination", NBER Technical Working Paper 192.

ECB (2002), "Monthly Bulletin March 2002", ECB.

Fildes, R. and Stekler, H. (2002), "The state of macroeconomic forecasting", Journal of Macroeconomics 24, pp. 435–468.

Fritzer, F., Moser, G. and Scharler, J. (2002), "Forecasting Austrian HICP and its components using VAR and ARIMA models", OeNB Working Paper 73.

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

Harvey, D., Leybourne, S. and Newbold, P. (1998), "Tests for forecast encompassing", Journal of Business and Economic Statistics 16, pp. 254–259.

Hubrich, K. (2003), "Forecasting euro area inflation: Does contemporaneous aggregation improve the forecasting performance?", ECB Working Paper 247.

Kunst, R. (2003),"Testing for relative predictive accuracy: A critical viewpoint", Institute for Advanced Studies, Vienna, Economics Series 130.

Marcellino, M., Stock, J. H. and Watson, M. W. (2000), "Macroeconomic forecasting in the euro area: country specific versus area-wide information", IGIER Working Paper 201.

Stock, J. H. and Watson, M. W. (1999a), "Diffusion indexes", Princeton University, NBER Working Paper W6702.

Stock, J. H. and Watson, M. W. (1999b), "Forecasting inflation", Journal of Monetary Economics 44, pp. 293–335.

Tkacz, G. and Gosselin, M. A. (2001), "Evaluating factor models: an application to forecasting inflation in Canada", Working Paper 2001–18, Bank of Canada.

# Appendix

## A. "Backcasting" the HICP Using the CPI

The HICP and its sub-indices are available from Statistik Austria beginning in January 1987. As noted by Ashley (2003), a sufficient number of observations in the validation period, preferably above 100, is necessary to establish significant differences in predictive accuracy. This implies that estimation has to start at an earlier date than 1987. The problem therefore is to "backcast" the HICP and its sub-indices.

We chose the following approach: First, based on qualitative information for each subindex and the HICP, those CPI sub-indices are identified that are related to the corresponding HICP indices. Then these indices are used together with all other available CPI indices in the regression $\Delta P_{i,t} = c + \sum_{j=0}^{N} \theta_j \Delta CPI_{j,t} + \varepsilon_{i,t}$. In that regression the annual increase of the HICP or HICP subindex, $\Delta P_{i,t}$, is regressed on the annual increases of $N$ sub-indices of the CPI and the CPI itself and a constant. The next step is to exclude sequentially all CPI components that are not significant, except those that are already identified to be related to the HICP or a subindex of the HICP based on qualitative information. The result is a parsimonious representation of the HICP or HICP subindex in terms of the CPI and/or CPI sub-indices. This procedure reflects the hypothesis that the set of prices of individual goods contained in the HICP and CPI is largely similar but aggregated in different fashions.

The results of these regressions (available from the authors on request) display several common features: First, the adjusted $R^2$ is very high, between 88 and 96%. Second, in all equations there is considerable serial correlation, with DW-statistics ranging from 0.32 to 0.71. The good fit is evidence that the HICP sub-indices are well approximated by the CPI sub-indices. However, the low DW-statistic in conjunction with the high $R^2$ points to a possible spurious relationship. The next step consists of using the estimated coefficients $\hat{\theta}_j$ to generate predicted values for the HICP and the HICP sub-indices for the period from January 1980 to December 1986. This increases the number of observations by 43%.

The expanded series are then subjected to two quality checks: First, a visual inspection of the series does not suggest the presence of major breaks in January 1987. Furthermore, for some indices also the seasonal pattern is clearly maintained. Second, the HICP sub-indices predicted by the above models are aggregated using the HICP weights of January 1987. The resulting backcasted aggregated HICP is

compared to the backcasted aggregate HICP. Here it turns out that the discrepancy between the two series in the pre-1987 period is comparable to this discrepancy in the post-1987 period (mentioned in footnote 9). Overall, it seems that the approximation of the HICP and the HICP sub-indices works fairly well. This suggests that the costs given by any approximation errors are outweighed by the gains in terms of increased discriminatory power across forecasting models mentioned above.

## B. Removing Breaks from the Price Index Series

A visual inspection of the monthly differences and the level of the seasonally adjusted price index series of the processed food and the industrial goods price index reveal that both series contain breaks. The processed food index displays an upward jump in the price level in January and February 1992 and a sharp fall in the price level in January and February 1995. The sizes of these shifts suggest that some event caused a temporary increase in the price level of processed food in Austria in the period from 1992 to 1994, followed by a permanent fall since then. A similar pattern is visible in the wholesale price index of food, but not in the index for unprocessed food or world market prices of food measured by the corresponding HWWA index. This points to a sectoral cause, possibly related to Austria's accession to the EU in January 1995. In order to remove the two breaks from the series a dummy variable is defined as shown in chart 1, which is then subtracted from the original series to obtain the adjusted series as displayed in chart 2.

*Chart 1: Dummy for the Shift in the Processed Food Index*[15]

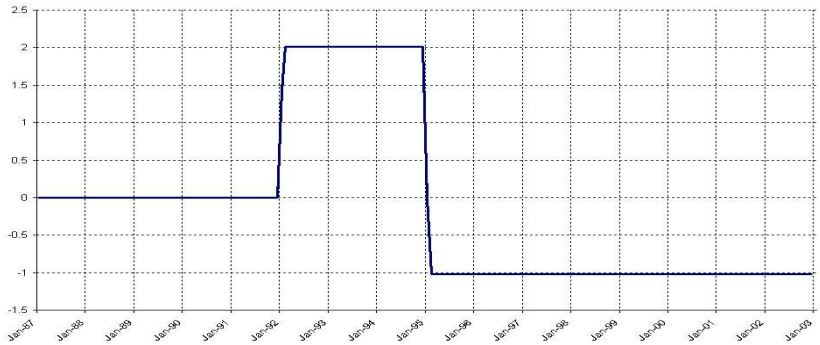
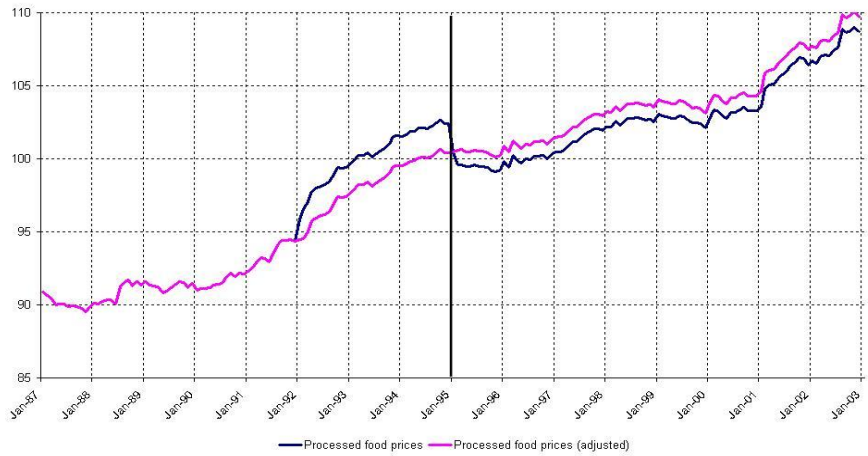



*Chart 2: Processed Food Price Level*




[15] All data used in charts originate from authors' calculations.

*Chart 3: Dummy Variable for the Industrial Goods Index*

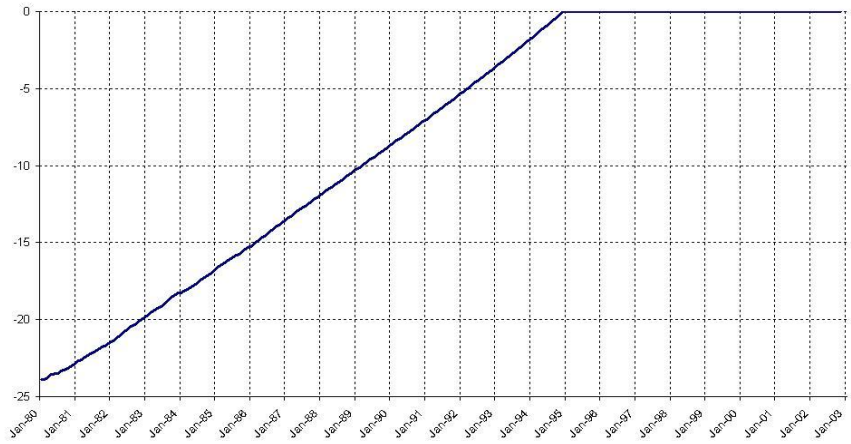
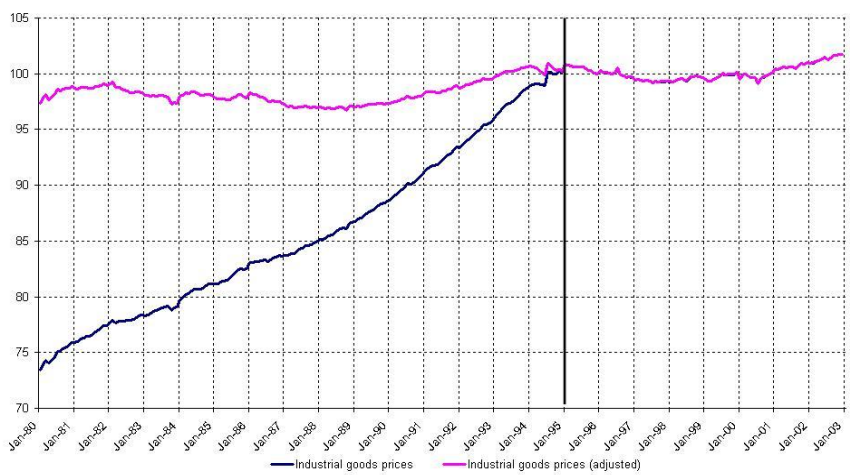



*Chart 4: Industrial Goods Price Level*



The level and monthly differences of the industrial goods index suggest a break in January 1995, possibly related to increased competitive pressure induced by EU entry. The break is dealt with by calculating the mean growth rates for the pre- and post-1995 periods and subtract the difference between the two means from the pre-1995 data. This removes the strong upward trend visible in the pre-1995 data (see charts 3 and 4). The remaining variation in the manipulated series reflects short run changes in inflation, the focus of our forecasting efforts.

The headline HICP is also affected by the two breaks. They are removed from the index by multiplying the dummies for the two indices with the weights of these indices in the HICP and subtracting it from the HICP.

## C. Test Statistics for Comparing Predictive Accuracy and Encompassing

Calculating the modified Diebold-Mariano statistic of Harvey, Leybourne and Newbold proceeds as follows. A loss differential sequence, $d_{1,j,t+12} = \hat{u}^2_{j,t+12} - \hat{u}^2_{1,t+12}$ is computed where $\hat{u}_{1,t+12}$ and $\hat{u}_{j,t+12}$ denote the 12-step out-of-sample forecast errors of the factor model forecasts and the forecasts of the competing model. The mean of this sequence is given by $\bar{d}_{1,j} = N^{-1}\sum_{t=1}^{N} d_{1,j,t+12}$. Note that $\bar{d}_{1,j} = MSE_j - MSE_1$. Furthermore, let $\hat{S}^{DM^{mod}}_{1,j} = \gamma_{1,j,0} + 2\sum_{k=1}^{n} \gamma_{1,j,k}$ be a consistent estimate of the long-run covariance of the sequence $d_{1,j,t+12}$, where $\gamma_{1,j,k} = N^{-1}\sum_{t=1}^{N} (d_{1,j,t+12} - \bar{d}_{1,j})(d_{1,j,t+12-k} - \bar{d}_{1,j})$. Following Diebold and Mariano, our choice of the truncation lag $n$ is motivated by the fact that optimal k-step forecasts should display at most $(k-1)$-dependence. Since our forecasts are 12-step we choose a truncation lag of 14 to account for deviations from optimality. Only those autocovariances of the sequence $d_{1,j,t+12}$ enter the long-run covariance with a non-zero value which are at lags with a significant autocorrelation coefficient. Significance is given if the absolute value of an autocorrelation coefficient is greater than $2/\sqrt{N}$. Note that $d_{1,j,t+12}$ has to be stationary which is checked using the augmented Dickey-Fuller test. The test statistic

$$DM^{mod}_{1,j} = \left[\frac{N+1-2k+N^{-1}k(k-1)}{N}\right]^{0.5} \frac{\bar{d}_{1,j}}{\sqrt{\hat{S}^{DM^{mod}}_{1,j}/N}} \tag{9}$$

is given by the difference between the mean squared errors of the two models, scaled by the standard deviation of the sequence $d_{1,j,t+12}$. The expression in square brackets is the size correction proposed by Harvey, Leybourne and Newbold. They note that this test still has the tendency to reject a true null somewhat to often. The measure of the standard deviation accounts for the autocorrelation in the loss differential sequence which may be present due to our multi-step forecasting framework.

Calculating the encompassing test statistic of Harvey, Leybourne and Newbold proceeds as follows. Let $\hat{c}_{s,j,t+12} = \hat{u}^2_{s,t+12} - \hat{u}_{j,t+12}\hat{u}_{s,t+12}$ and $\bar{c}_{s,j} = N^{-1}\sum_{t=1}^{N} \hat{c}_{s,j,t+12}$. The index $s$ denotes the forecast with higher predictive accuracy as established by the modified Diebold Mariano test and/or the descriptive statistics and $j$ denotes the competing forecast. Note that $\bar{c}_{s,j} = MSE_s - cov(\hat{u}_{j,t+12}, \hat{u}_{s,t+12})$. The statistic is given by

$$HLN_{s,j} = \frac{\bar{c}_{s,j}}{\sqrt{\hat{S}^{ENC}_{s,j}/N}}, \tag{10}$$

where $\hat{S}^{ENC}_{s,j} = \delta_{s,j,0} + 2\sum_{k=1}^{n} \delta_{s,j,k}$ and $\delta_{s,j,k} = N^{-1}\sum_{t=1}^{N} (\hat{c}_{s,j,t+12} - \bar{c}_{s,j})$ $(\hat{c}_{s,j,t+12-k} - \bar{c}_{s,j})$. The truncation lag $n$ is again set to 14. Under the null hypothesis that the forecast with higher predictive accuracy encompasses the forecast of the competing model, the difference between the MSE of the model with higher predictive accuracy and the covariance between the forecast errors will be less than or equal to zero. Under the alternative that the competing model contains additional information, the difference should be positive and large compared to the standard deviation of the sequence $\hat{c}_{s,j,t+12}$. This condition is more likely to be fulfilled if the forecast errors of the two models are negatively correlated.

# D. List of Data

Labor market

1. Unemployment, total
2. Unemployment, female
3. Unemployment, male
4. Unemployment, construction sector
5. Unemployment rate, total
6. Unemployment rate, female
7. Unemployment rate, male
8. Employment, total
9. Employment, female
10. Employment, male
11. Employment, total, blue collar
12. Employment, female, blue collar
13. Employment, male, blue collar
14. Employment, total, white collar
15. Employment, male, white collar
16. Employment, female, white collar
17. Employment, foreigners
18. Vacancies

Trade balance

1. Imports, food
2. Imports, raw materials
3. Imports, intermediate goods
4. Imports, finished goods
5. Imports, finished goods, investment goods
6. Imports, finished goods, consumption goods
7. Imports, finished goods, miscellaneous
8. Imports, machinery, vehicles
9. Imports, total, excluding intra euro area dispatches
10. Imports, total
11. Imports, total, unit values
12. Exports, food
13. Exports, raw materials
14. Exports, intermediate goods
15. Exports, finished goods
16. Exports, finished goods, investment goods
17. Exports, finished goods, consumption goods
18. Exports, finished goods, misc.

19. Exports, machinery, vehicles
20. Exports, total, excluding intra euro area dispatches
21. Exports, total
22. Exports, total, unit values

Money and credit

1. Deposits with maturity up to two years
2. Demand deposits
3. M1
4. M2
5. M3
6. Loans to the private sector
7. Collateralized loans
8. Foreign currency loans
9. Private sector demand deposits
10. Private sector time deposits
11. Cash in stock at banks
12. Deposits of banks at central bank
13. Liquidity of banks

Wholesale prices

1. Wholesale prices (total)
2. Consumer goods (total)
3. Consumer goods (durable)
4. Consumer goods (non-durable)
5. Consumption goods
6. Intermediate goods
7. Construction goods
8. Investment goods
9. Iron and steel
10. Non-steel metals
11. Solid Fuels
12. Food
13. Electrical appliances
14. Paper and paper products
15. Seasonal food
16. Feed barley
17. Soy grits
18. Utility calfs
19. Calf breed
20. Chicken
21. Pork chop

22. Beef
23. Veal
24. Pulp wood (Styria)
25. Pulp wood (Upper Austria)
26. Energy

Aggregate demand

1. Industrial production (total)
2. Industrial production excluding energy and construction
3. Industrial orders
4. Industrial sales price expectations
5. Car registration and sales

Negotiated monthly wages

1. All employees, total,
2. All employees, excluding public services
3. All employees, public services
4. All employees, public services, transportation
5. All employees, industry
6. All employees, manufacturing
7. All employees, construction
8. All employees, trade
9. All employees, transportation
10. All employees, tourism
11. All employees, agriculture and forestry
12. Blue collar, total
13. Blue collar, industry
14. Blue collar, construction
15. Blue collar, manufacturing
16. Blue collar, trade
17. Blue collar, transportation
18. Blue collar, tourism
19 Blue collar, agriculture and forestry
20 White collar, total
21. White collar, industry
22. White collar, construction
23. White collar, manufacturing
24. White collar, trade
25. White collar, transportation
26. White collar, tourism
27. White collar, banking
28. White collar, agriculture and forestry

Raw materials

1. Import prices, coal
2. Import prices, electricity
3. Import prices, crude oil, including components for processing
4. Import prices, crude oil
5. Import prices, liquid gas
6. Import prices, gasoline
7. Import prices, heating oil
8. HWWA index, total
9. HWWA index, total, excluding energy
10. HWWA index, food and tobacco
11. HWWA index, materials used in manufacturing
12. HWWA index, materials used in agriculture
13. HWWA index, non-steel metals
14. HWWA index, iron ore and scrap
15. HWWA index, energy
16. HWWA index, coal
17. HWWA index, crude oil
18. Brent crude oil

Tourism

1. Price index for foreigners in Austria
2. Price index for domestic residents in foreign countries
3. Price index for domestic residents in Austria
4. Bednights, total
5. Foreign tourist bednights
6. Domestic tourist bednights

Exchange rates

1. Austrian Schilling to the U.S. dollar
2. Austrian Schilling to the Canadian dollar
3. Austrian Schilling to the pound sterling
4. Austrian Schilling to the Swiss francs
5. Austrian Schilling to the Norwegian krone
6. Austrian Schilling to the Swedish krone
7. Austrian Schilling to the Japanese yen
8. Austrian Schilling to the Australian dollar
9. Austrian Schilling to the Korean won
10. Austrian Schilling to the Indonesian rupiah
11. Austrian Schilling to the Thai baht
12. Austrian Schilling to the Malaysian ringgit
13. Austrian Schilling to the Philippine peso

14. Effective exchange rate, nominal
15. Effective exchange rate, real, CPI based
16. Effective exchange rate, real, Ulc-mfg based
17. Terms of trade index, domestic currency

Interest rates

1. Yield on German government bond, one year residual maturity
2. Yield on German government bond, two years residual maturity
3. Yield on German government bond, three years residual maturity
4. Yield on German government bond, four years residual maturity
5. Yield on German government bond, five years residual maturity
6. Yield on German government bond, six years residual maturity
7. Yield on German government bond, seven years residual maturity
8. Yield on German government bond, eight years residual maturity
9. Yield on German government bond, nine years residual maturity
10. Yield on German government bond, ten years residual maturity
11. Base rate of the Oesterreichische Nationalbank
12. Reference rate of the Oesterreichische Nationalbank
13. Yield on Austrian government bond, ten years residual maturity
14. Vienna stock exchange, index
15. Overnight interest rate, Frankfurt
16. Three months deposit interest rate, Zurich
17. Three months deposit, Eurodollar
18. Federal Funds rate
19. Yield, secondary market government bonds, bond basket, other maturities.
20. Yield, private sector bonds, including bank and mortgage bonds.
21. Interest rate, euro-currency, 1-month bid rate
22. Interest rate, euro-currency, 3-month bid rate
23. Interest rate, euro-currency, 6-month bid rate
24. Interest rate, euro-currency, 12-month bid rate
25. 3-month VIBOR

CPI and CPI components

1. CPI
2. CPI excluding unprocessed food
3. Food and beverages
4. Unprocessed food
5. Food
6. Services
7. Rents
8. Tobacco
9. Rents and maintenance of flats

10. Lighting und heating
11. Clothing and personal equipment
12. Cleaning of clothing
13. Personal hygiene
14. Leisure and education
15. Transport

HICP and HICP components

1. HICP
2. Processed food
3. Unprocessed food
4. Energy
5. Industrial goods
6. Services

*Chart 5: Acf of the HICP Inflation Forecast Error*

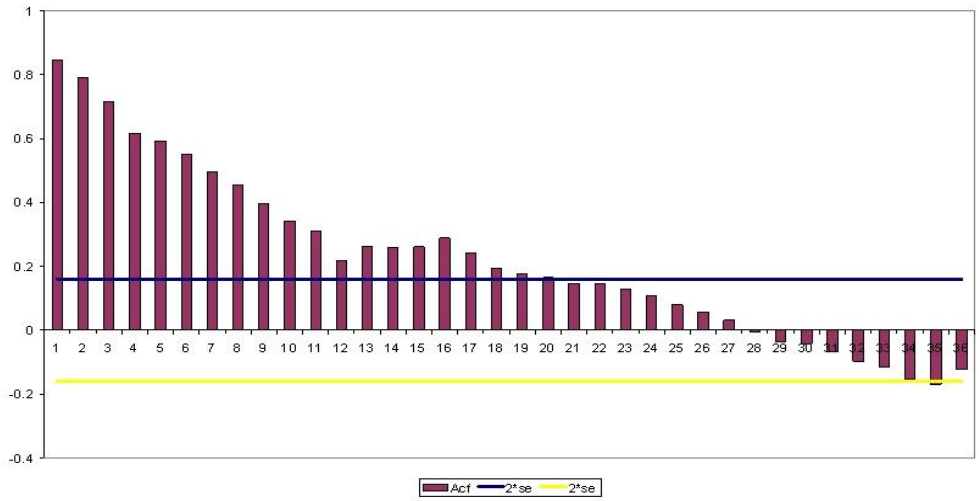



*Chart 6: Acf of the Processed Food Price Inflation Forecast Error*

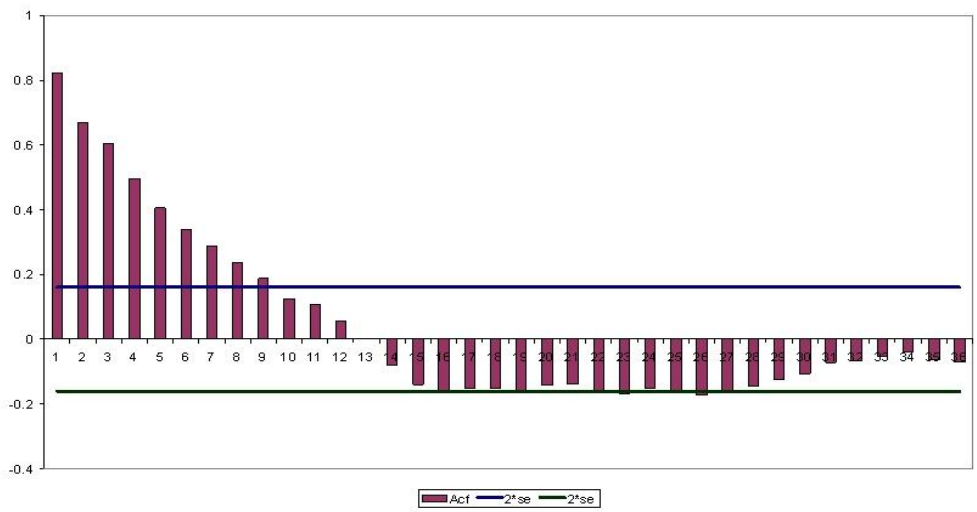


Chart 7: Acf of the Unprocessed Food Price Inflation Forecast Error

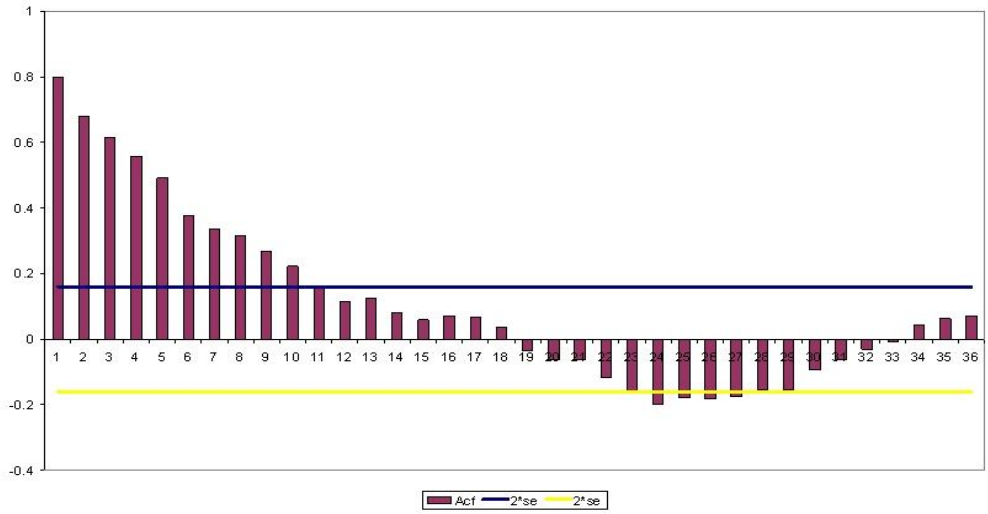



*Chart 8: Acf of the Energy Price Inflation Forecast Error*

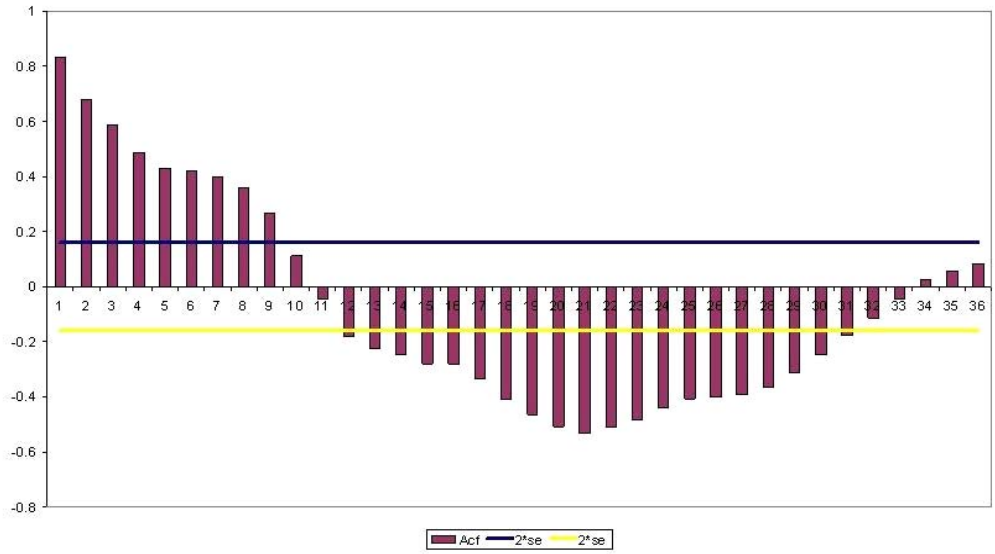


*Chart 9: Acf of the Industrial Goods Price Inflation Forecast Error*

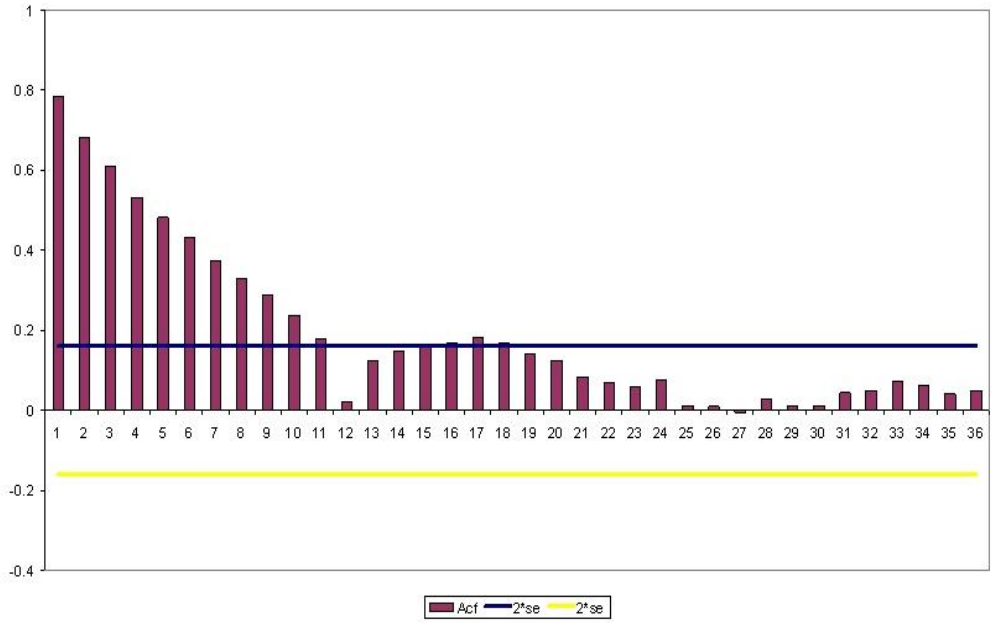



*Chart 10: Acf of the Services Price Inflation Forecast Error*

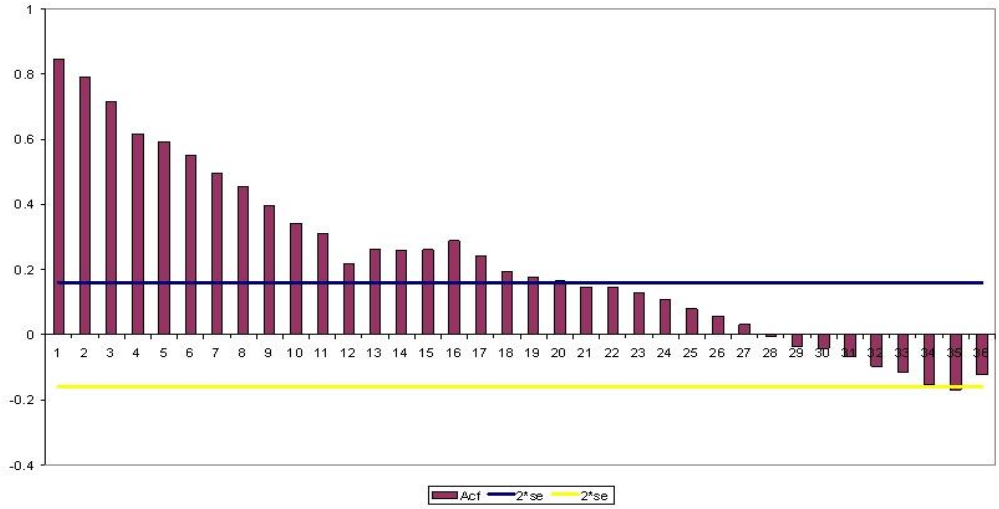


*Chart 11: Forecast Error – HICP*

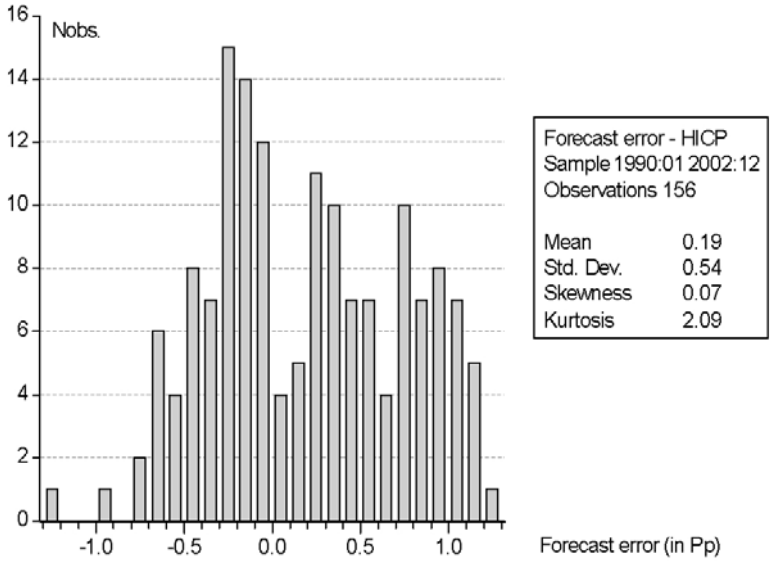



*Chart 12: Forecast Error – Processed Food*

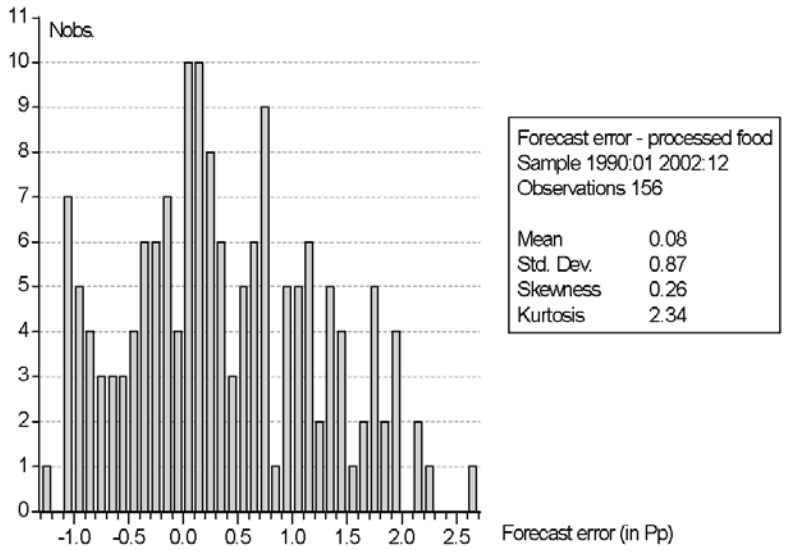


*Chart 13: Forecast Error – Unprocessed Food*

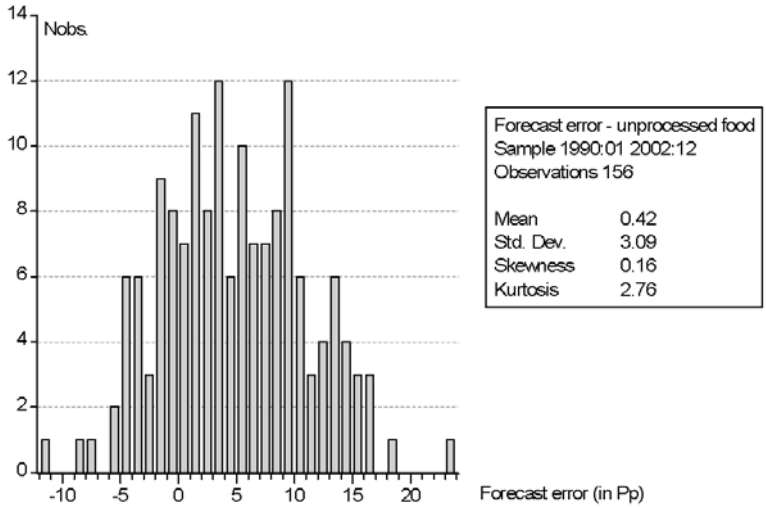



*Chart 14: Forecast Error – Energy*

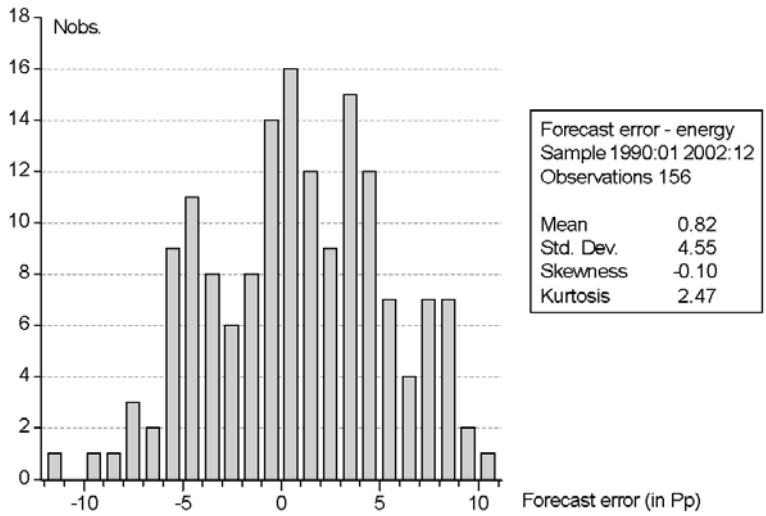


*Chart 15: Forecast Error – Industrial Goods*

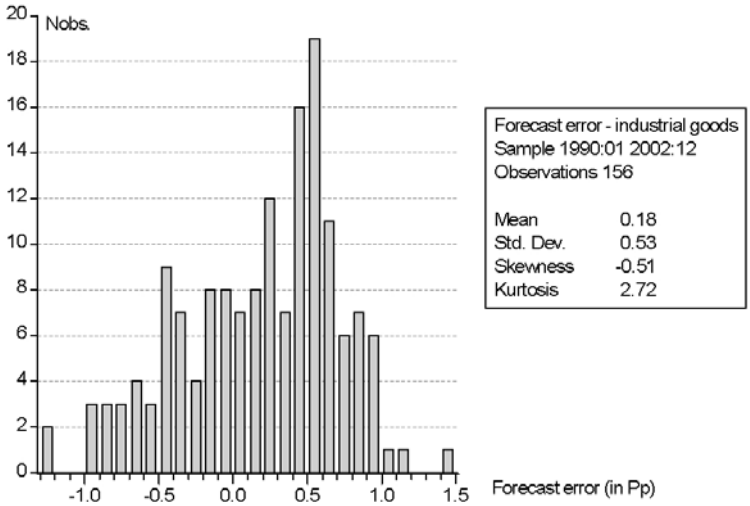



*Chart 16: Forecast Error – Services*

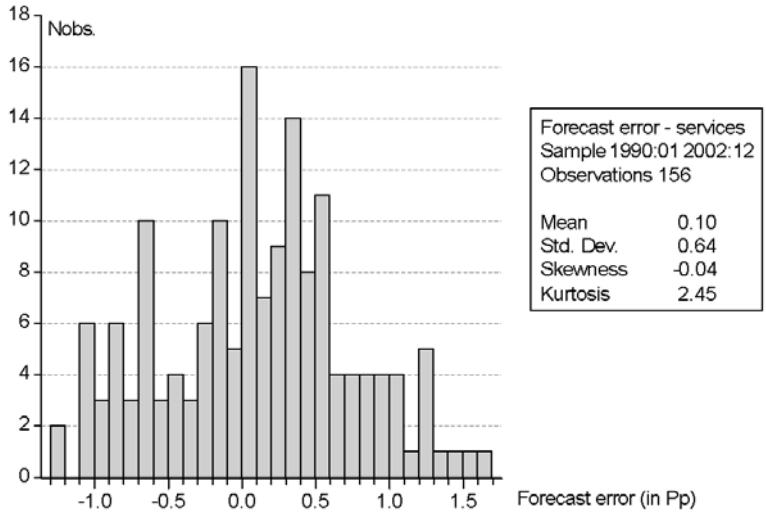


*Chart 17: HICP Inflation*

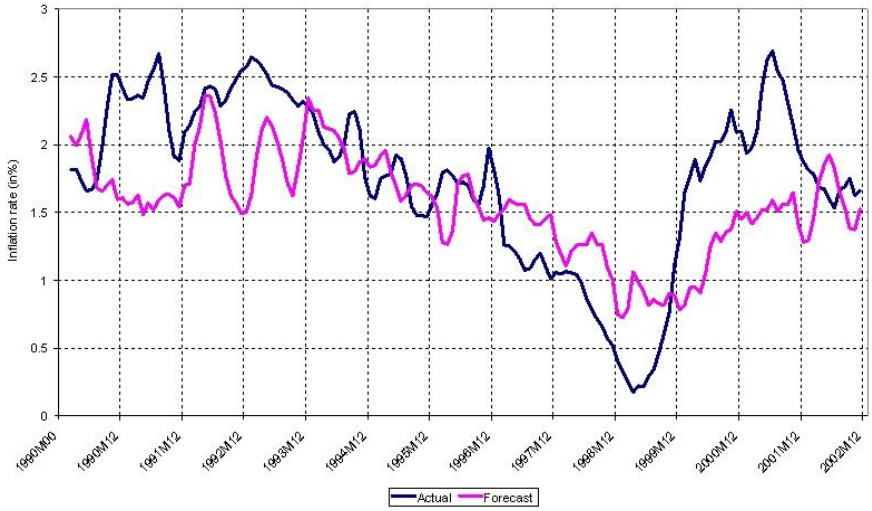



*Chart 18: Processed Food Price Inflation*

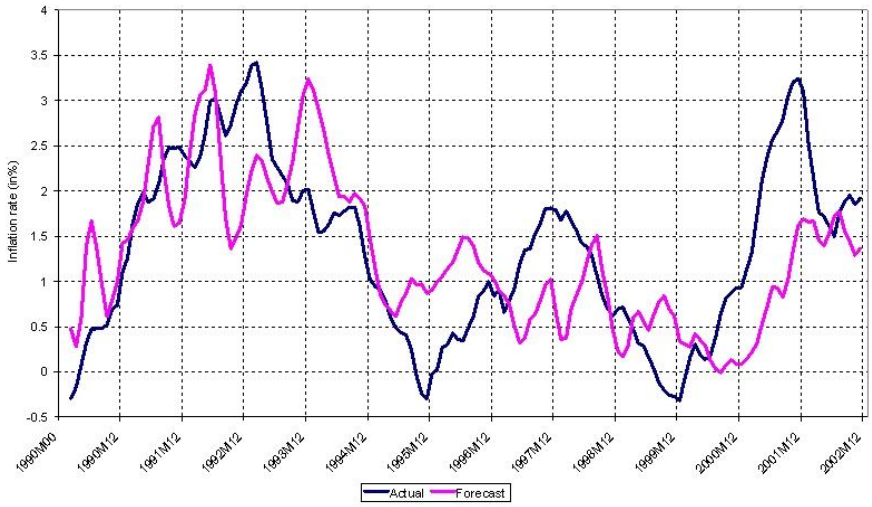


*Chart 19: Unprocessed Food Price Inflation*

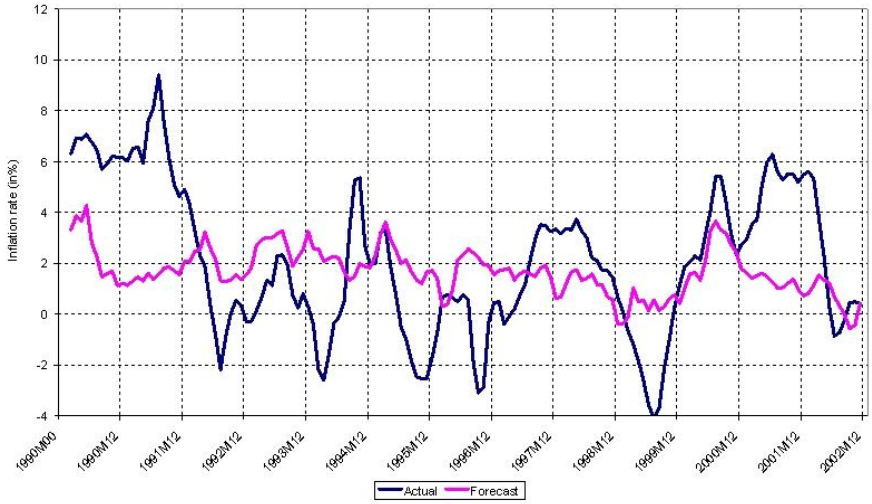



*Chart 20: Energy Price Inflation*

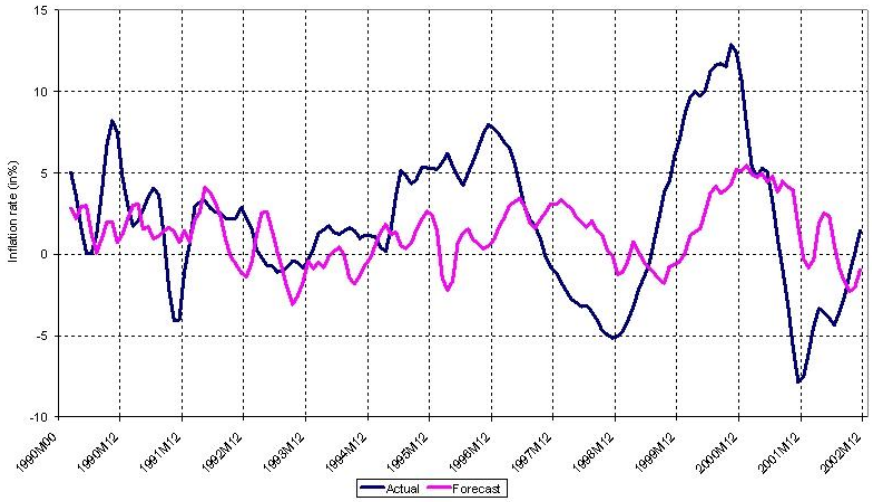


*Chart 21: Industrial Goods Price Inflation*

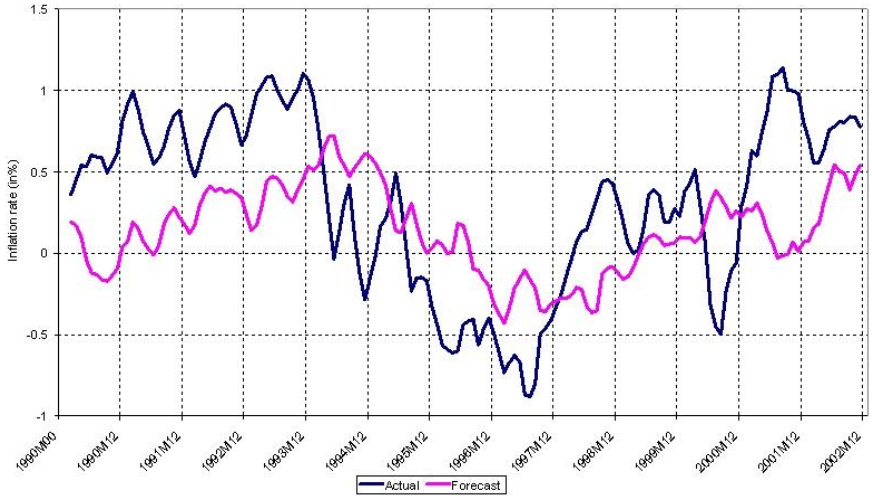



*Chart 22: Services Price Inflation*

*Chart 23: HICP–Forecast Error*

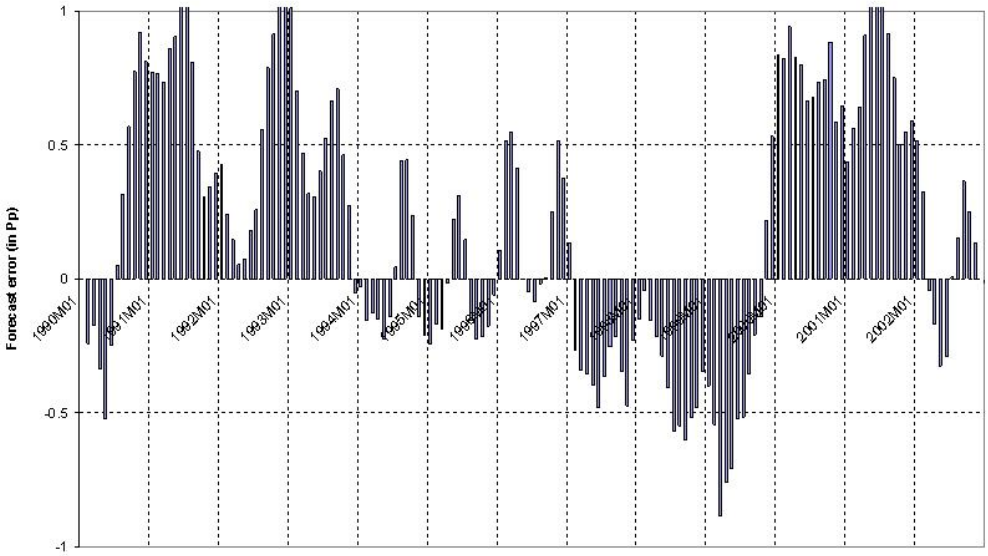



*Chart 24: Processed Food – Forecast Error*

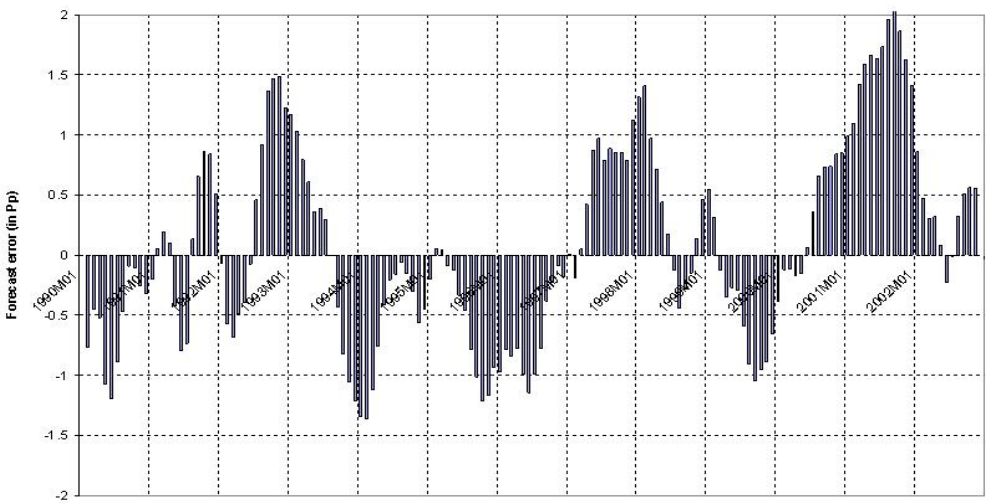


*Chart 25: Unprocessed Food – Forecast Error*

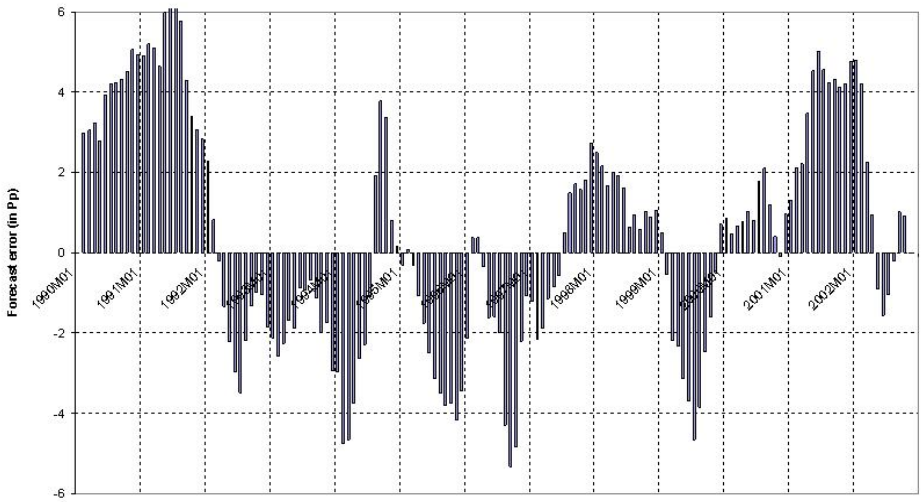



*Chart 26: Energy – Forecast Error*

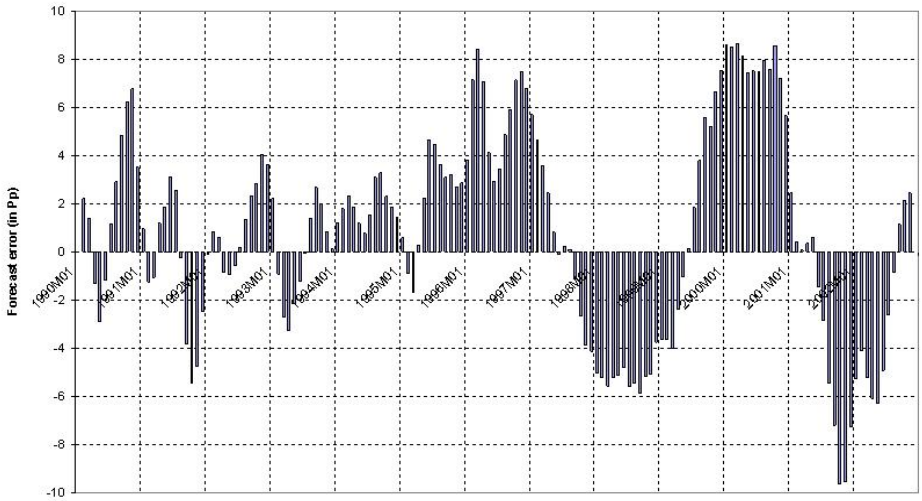


*Chart 27: Industrial Goods – Forecast Error*

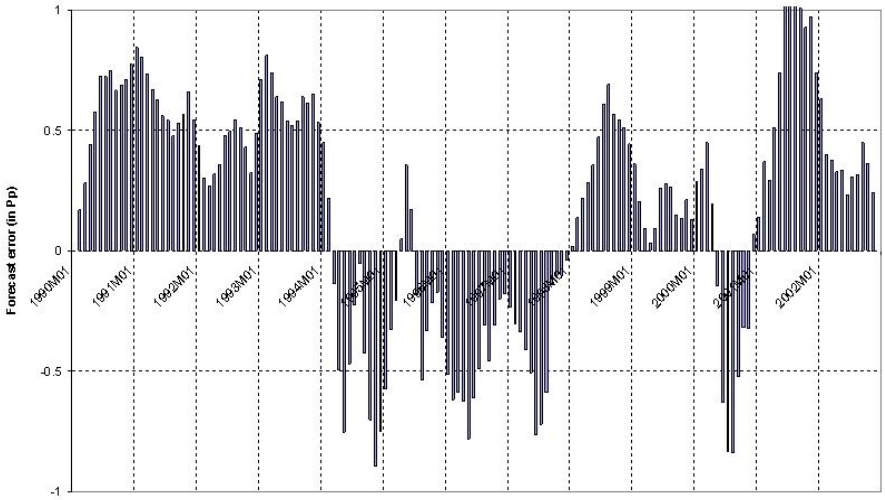



*Chart 28: Services – Forecast Error*

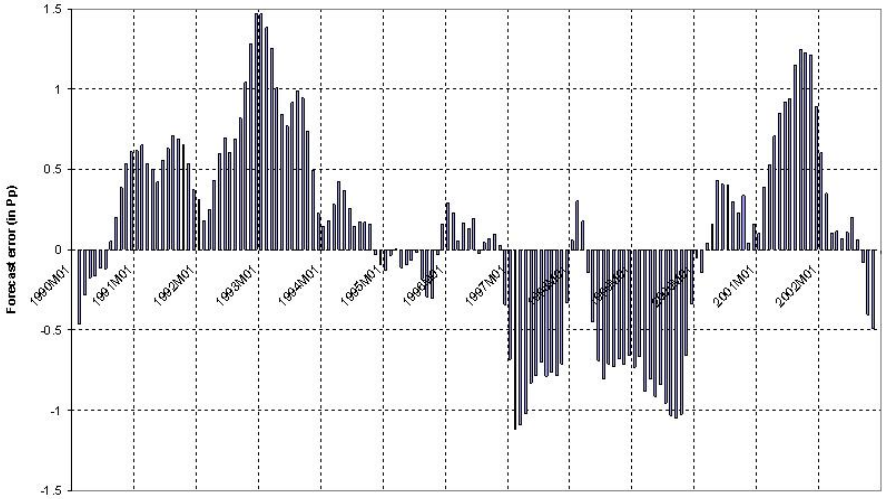


*Chart 29: HICP – Forecast Decomposition*

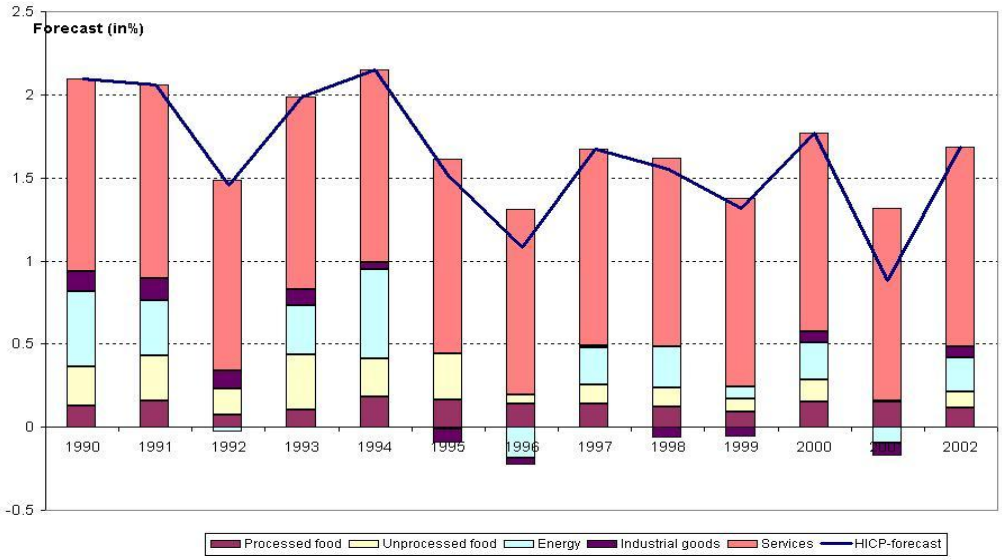



*Chart 30: HICP – Forecast Error Decomposition*