FORECASTING INFLATION: AN ART AS WELL AS A SCIENCE!***

BY

ARD DEN REIJER*,** AND PETER VLAAR**

Summary

In this study, we build two forecasting models to predict inflation Harmonised Index of Consumer Prices (HICP) for the Netherlands and for the euro area. The models provide point forecasts and prediction intervals for both the components of the HICP and the aggregated HICP -index itself. Both models are small-scale linear time series models allowing for long-run equilibrium relationships between HICP components and other variables, notably the hourly wage rate and the import or producer prices. The model for the Netherlands is used to generate the Dutch inflation projections for the eurosystem's Narrow Inflation Projection Exercise (NIPE). The recursive forecast errors for several forecast horizons are evaluated for all models, and are found to outperform a naive forecast and optimal AR models. Moreover, the same result holds for the Dutch NIPE projections, which have been provided quarterly since 1999. The aggregation method to predict total HICP inflation generally outperforms the direct method, except for long horizons in the case of the Netherlands.

Key words: aggregation, model selection time series models

JEL Code(s): C32, C43, C52, C53

1 INTRODUCTION

The mandate of the European Central Bank (ECB) is to maintain price stability in the euro area. This goal is given a quantitative content by requiring that the year on year growth of the Harmonised Index of Consumer Prices (HICP) for the euro area as a whole should be close to, but below 2% in the medium term. The ECB is monitoring and forecasting price developments under the first pillar of its monetary policy strategy.[1] Therefore, forecasting

* Corresponding author: a.h.j.den.reijer@dnb.nl

** De Nederlandsche Bank, Research Division, P.O. Box 98, 1000 AB Amsterdam, the Netherlands;

*** We thank an anonymous referee of this journal and participants of the second Conference of the Euro Area Business Cycle Network, the 10th International Conference on Computing in Economics and Finance and seminar participants at De Nederlandsche Bank, Erasmus University Rotterdam and Maastricht University, in particular Filippo Altissimo, Bob Chirinko, Bertrand Candelon, Denise Osborn and Adrian Pagan for useful comments.

1 In practice, each country provides four times a year its own inflation forecast for an horizon of 11–15 months and these forecasts are used to construct an area wide forecast. This periodic procedure is called the NIPE.

inflation rates has become important for both monetary authorities and private agents who try to understand and react to the central bank's behaviour. The aim of this paper is to describe the procedures used at De Nederlandsche Bank to predict Dutch HICP inflation and a new model to directly predict overall euro area inflation. The models forecast both the components of the HICP, as requested by the euro system's Narrow Inflation Projection Exercise (NIPE), and, for comparison reasons, the total HICP itself. The forecast horizon is 11–18 months ahead. Similar recent papers describing small-scale linear models aimed at forecasting inflation include Banerjee et al. (2003), Benalal et al. (2004), Bruneau et al. (2003a,b), Fritzer et al. (2002), Hubrich (2005) and Moser et al. (2004).

One of the issues addressed in this paper is that of disaggregation, which can be regarded in three dimensions, that is across components (sub-indices of an index), across time (higher frequency) and across space (different regions of an economic area). In the European context, there is the aggregation of the forecasts of individual countries to a euro area level; see for instance Espasa et al. (2002), Marcellino et al. (2003) and Benalal et al. (2004). In this study, we will only address the aggregation of HICP component forecasts.[2] Obviously, there is a clear interest in finding out whether aggregating component forecasts performs better than forecasting the aggregate directly. Aggregating forecasts of component models is potentially beneficial as forecast errors might cancel between components. Moreover, the disaggregate components can be better modelled by choosing a more suitable model for each component separately and by possibly incorporating additional explanatory variables. This argument is indeed apparent in theoretical models, see Lütkepohl (1987) and Hendry and Mizon (1999). The latter authors assume for instance a known and constant data generation process. However, Hubrich (2001, 2005) and Benalal et al. (2004) find empirical evidence for euro area data and across various specifications that directly forecasting the aggregate HICP performs better than aggregating the forecasted components, especially for a forecast horizon up to 12 months ahead. Apart from the possible efficiency gain, a policy maker also has an interest in forecasting the HICP components to construct a measure for core inflation, defined as HICP excluding the components unprocessed food and energy. These two components are generally considered more volatile and less susceptive to monetary policy.

Another important issue in this paper is the model selection procedure, for which a new heuristic method is developed. The method involves three steps. The first step involves the visual inspection of the data, primarily to detect changing seasonal patterns. This step determines the general structure of the

2 In contrast, the eurosystem's NIPE creates a euro area forecast by aggregating the individual countries' inflation forecasts. So, the model for the euro area built in this study generates forecasts for the area wide aggregates, while the NIPE aggregates the individual country's forecasts to the euro area level.

small-scale model. The second step involves calculating all possible models given this structure, allowing for a small set of exogenous and endogenous variables and variable lag lengths. Optimal statistical models are subsequently selected according to goodness-of-fit, parsimony and/or out-of-sample forecasting accuracy. The final selection is based on the economic evaluation of the statistically selected models. Especially, the long-run properties are important in this respect.

This paper is organized as follows. In section 2, we describe our procedure to select an optimal forecasting model. The selected models for the Netherlands and the euro area are described in section 3. Section 4 elaborates on the uncertainty surrounding the forecasts. In section 5, the forecast results of the models are evaluated. First, the recursive root mean squared forecast errors are compared to those of random walk models and optimal autoregressive models, both for the component models and the direct HICP models. Then, the Dutch NIPE results are evaluated. Finally, section 6 concludes.

2 MODEL SELECTION PROCEDURE

The model selection procedure consists of three steps. The first step involves preliminary data analyses. This step is used to select the optimal model structure. That is to say, either a vector autoregressive (VAR) model in first differences or a vector error correction model (VECM) in both first and 12 month differences. The second step involves the computation of all possible models given the possible set of exogenous and endogenous variables and the allowable lag length. The optimal models according to several statistical criteria are subsequently shown. The last step implies the economic evaluation of the statistically selected models in order to select the final optimal model. This involves both an economic interpretation of the coefficients, primarily with respect to the error correction term and an analysis of the model properties to provide stable long-run forecasts.

2.1 *Preliminary Data Analysis*

The main purpose of this first step is to detect time varying seasonal patterns as the HICP data are not seasonally adjusted. They are plotted in appendix, Figure A1-A2 for, respectively, the Netherlands and the euro area. All series in both cases are plotted in raw format, in annual inflation rates, that is 12 month differences of the log-transformed HICP and in the monthly change of the log price index, that is monthly inflation. The figures show that especially the (log) sub-indices for non-energy industrial goods (P^i) and services (P^s) as well as total HICP (P^{total}) are subject to a changing seasonal pattern. Particularly for the euro area but also for the Netherlands, this change is not concentrated in just 1 month. Otherwise, a second set of seasonal dummies should be sufficient to remove the seasonality. Instead, the pattern is filtered out of the data by taking 12 month differences. Moreover first differences are taken to eliminate the (near) unit root in inflation. However, lagged annual inflation is also included to prevent overdifferencing. In this way, stationarity of annual inflation, or cointegration of inflation with other variables is allowed for. The same model is used for processed food (P^{pf}) for the Netherlands, for which the seasonal pattern is less clear. Concerning the series unprocessed food (P^{uf}), energy (P^{e}), and P^{pf} for the euro zone, seasonal dummies are included in the model to capture the seasonal effects. As these variables are clearly non-stationary, they are modelled in first differences. Cointegration is not allowed for these models in first differences as it appears that the economic rationale for a long-run relationship among price levels is less obvious than among inflation rates. Moreover, even if such a relationship would exist, it is prone to structural breaks due to for instance indirect tax adjustments.

2.2 *Statistical Criteria*

Given the model structure dictated by the seasonal pattern, all possible models given the set of explanatory variables are computed in an automated selection procedure.[3] That is to say, a model is computed for every possible combination of explanatory variables and every possible lag structure from zero up to a maximum (usually 12).[4] Moreover, the 12th lag is analysed separately as this lag is theoretically important for monthly data. For instance, if inflation is surprisingly low in a certain month due to an earlier start of summer sales, it can be expected that inflation will be relatively high the next year in the same month (unless the change in sales pattern is permanent). For models in first and 12th differences, a negative AR(12) coefficient can also partly remedy the permanent base effects of, for instance, indirect tax changes. The same lag structure is used for all the variables included. Moreover, for the VECMs, every possible combination of lagged 12 month differences including inflation itself is added to check for long-run relationships (and to correct for possibly incorrectly imposed unit roots in inflation).[5]

3 The forecasts of natural gas prices (part of P^{e}) and housing rents (part of P^{s}) for the Netherlands are generated outside the model as they are for the time being only adjusted twice, respectively, once a year, according to some strict rules.

4 However, some limits are imposed on the total number of variables and/or parameters in the model in order to preserve enough degrees of freedom.

5 The selection of the long-run relationship is based on the same criteria as the variable selection, the lag length and the inclusion of the autoregressive term at lag 12. No formal cointegration rank tests are performed. As the left-hand side variable (the monthly change in inflation) is clearly stationary, the 12 month differences will only enter if they are indeed cointegrated or stationary themselves.

Regarding the explanatory variables, both endogenous and exogenous explanatory variables motivated by economic theory and data availability are potentially included. Some exogenous variables are provided by the ECB for the NIPE-exercise practice, like the future paths for policy variables and the variables regarding the external environment of the EMU.[6] Besides these variables nominal wages are also included exogenously since overlapping contracts and other institutional features make them relatively easy to predict in the short run. Potential endogenous variables include producer prices, import prices, the nominal money stock, industrial production, credit data, retail turnover and business cycle indicators.

The primary selection criteria for the optimal models are modified versions of the classical information criteria existing in the VAR literature: Schwarz (*SC*), Hannan-Quinn (*HQ*) and Akaike (*AIC*). These standard criteria are primarily used to determine the optimal lag length of a given VAR system. They need to be modified for our goal as not only the lag length needs to be decided, but also the choice of additional (endogenous or exogenous) variables. Different models can hardly be evaluated on the basis of covariance matrices if models with different endogenous variables are compared, as the residuals of different equations are computed. Therefore, we calculate the information criteria based only on the number of parameters and the residuals of the inflation equation alone. A good fit for the inflation equation alone is not enough however, to guarantee a reasonable forecast. If the model includes other endogenous variables that can hardly be forecasted themselves, the inflation forecast will be hampered as well. Consequently, we also apply an alternative measure of fit widely used within the forecasting literature, namely the root mean squared forecast error. Too much focus on the out-of-sample performance on the other hand would favour exogenous variables too much as they are included assuming perfect foresight.[7] Moreover, given our relatively short sample, it seems hardly efficient to ignore the in-sample fit.[8]

The modified information criteria are therefore, based on the weighted average of the in-sample variance of the inflation equation and the out-of-sample forecast error variance; SC^{mixed}, HQ^{mixed}, AIC^{mixed}. Besides these mixed criteria, the in-sample criteria (SC^{in}, HQ^{in} and AIC^{in}) are also compared as well as the root mean squared (forecast) errors both in-sample

6 With respect to interest rates and exchange rates a no-change path is implemented, whereas futures are used to project oil and other commodity prices.

7 Including those variables endogenously instead is not an option as either the forecast is assumed to be conditional on the exogenous variables (provided by the ECB), or because they are included exogenously precisely because they are better forecasted using institutional knowledge than with a statistical model (wages).

8 Moreover, a purely out-of-sample selection method would select overparameterised models with positive probability (Inoue and Kilian (2005)).

(RMSFE^{in})[9], out-of-sample (RMSFE^{out}) and combined (RMSFE^{mixed}). These nine different criteria often provide nine different optimal models. In principal, the model selected by AIC^{mixed} is chosen. The relatively low penalty for extra parameters for this criterion is justified as the risk of overfitting is mitigated by the inclusion of the out-of-sample forecast variance. The other models might give important indications with respect to the preferred specification as well, however. The fact that the model selection choice is not robust with respect to the selection criterion puts some doubts on the existence of *the* optimal model. This is further confirmed by a periodic evaluation of the results. One more year of data often leads to different selected models.

2.3 *Economic Evaluation*

Given that different criteria prefer different models and the fact that these choices are not very robust with respect to the addition of more data, it is obvious that statistical criteria alone are hardly sufficient to select the optimal model. Judgmental issues, based on economic criteria are important as well. Here, three issues come to mind. First, does the choice of variables make sense? Second, are the parameter values in the model of the right sign and order of magnitude? Third, does the model include a stable anchor for long-run forecasts?

With respect to the undesirable variable selection, the example for services inflation in the Netherlands is typical. Both the within sample and the mixed criteria selected the oil price as an important variable for services inflation. In the estimated models, this was reflected in a very significant negative contemporaneous coefficient. As there is no economic rational for such a negative impact, the oil price was not included in our preferred model. The significance was probably due to an incidental correlation of outliers in the past. Indeed, according to model selection criteria based on the current data set, the oil price would no longer be selected by the mixed criteria, but the within sample criteria would still select it. Another example is the short-term interest rate. This variable showed up in the estimated models with a significant positive contemporaneous coefficient. Due to the widely acknowledged lag in monetary transmission, the central bank's interest rate actions to fight inflation apparently created a positive short-term relation between nominal and real interest rates and inflation.

In addition to checking the correct sign of the parameters, most attention regarding the order of magnitude of the coefficients is concentrated on the error correction term. In the automatic model selection procedure, all variables in the error correction term are already checked for their sign. Apart

9 The in-sample variance is hereby computed with a correction for the number of estimated parameters in order to get an unbiased estimate. Otherwise extra variables can only improve the result.

from that, implausible long-run elasticities might be remedied by slightly adjusting the model (for instance changing the lag length).

Probably, the most important economic criterion in the evaluation of models, is the presence of a stable conditional anchor for long-run forecasts. Here, the difference between endogenous and exogenous variables is essential. Including endogenous variables that are themselves hard to predict might lead to the drifting of inflation forecasts to unlikely regions, especially if this endogenous variable is included in the error correction term. Therefore for instance, the variable $M3$ is not allowed to appear in the cointegration relationship. This endogenous explanatory variable is very difficult to forecast over a longer horizon in this VECM setting, and a bad forecast would imply a severe bias in the long-run forecast of inflation. Christoffersen and Diebold (1998) show that error correction terms among endogenous variables alone do not help to produce better long-run forecasts as they have expectation zero in the long-run. Exogenous variables in an error correction term on the other hand do positively affect long-run forecasting as they steer the long-run outcome for the endogenous variables. Therefore, the wage development as an exogenous explanatory variable is imposed in the error correction term of the selected model if validated by the data. Wage development is well exogenously predictable due to sluggishness in the wage formation process and can act as an anchor of the model. Another anchor is formed by import prices although they are endogenous. However, the import prices themselves are well predictable from (exogenous) oil price and exchange rate developments.

3 THE EMPIRICALLY OPTIMAL MODELS

We applied the selection criteria on a sample running from 1987(10) and 1990(1) until 2002(8), respectively, for the Netherlands and the euro area. The in-sample errors are calculated from the model based on the sample up until 2000(12) and the forecasting errors are obtained from the sample 2001(1) onwards. The model selection is based on the number of 20 forecasts, which are generated using the realised values for the exogenous variables. The forecast errors of the exogenous variables are therefore excluded in the selection process, since the aim of the models is to produce inflation forecasts conditional on the exogenous explanatory variables.[10] As stated before, apart from the components P^{uf}, P^{e} as well as P^{pf} for the euro area, all models are specified in changes of 12 month differences. So, for most models this differencing implies a loss of 13 observations and a remaining sample size of $T = 166$ and $T = 139$ for, respectively, the Netherlands and the euro area. The sample for fitting the model is much larger than the sample for obtaining the out-of-sample forecast residuals. Although these forecast errors are less numerous,

10 For this reason, the Eurosystem's BMPE uses the word 'projection' to indicate that the forecast is actually conditional on exogenous assumptions.

they get a weight of 0.4 for the mixed criteria, so as to emphasise the importance of good forecasting performance.

We find in accordance with the literature, Stock and Watson (2003), that the forecasting track record of specific models and leading indicators is not invariant over time. Checking the robustness of the model specification by evaluating out-of-sample forecasts for different time periods is further complicated by small sample availability. Moreover, the different criteria produce different optimal models. Therefore, the model selection procedure is rerun regularly.

The selected models for the five components and the HICP-index are presented in Table 1 for the Netherlands and in Table 2 for the euro area. The optimal model for unprocessed food turns out to be a univariate random walk for the Netherlands and an AR(1) process for the euro area (both including seasonal dummies). Energy prices depend mainly on oil prices, and in the euro area also on producer prices. The most dominating explanatory variable for the other sub-indices is the wage rate, which is imposed and statistically validated in all cointegration relationships for both the Netherlands and the euro area. Wages tend to be more important for the Netherlands than for the euro area as revealed by the twice as high-Dutchlong-term coefficients. Besides wages, a relatively dominating leading indicator for the Netherlands is the import price index of Germany showing up in the cointegration relationship for all four indices.[11] For the euro area on the other hand, the producer price index is taking this role. These endogenous variables are themselves primarily driven by the oil price, the Euro/Dollar exchange rate and the commodity prices excluding energy. Finally in the euro area, unprocessed food inflation appears as explanatory variable in the index for processed food. This index of processed food turns out to be important for services, which can be explained by restaurant prices. A cointegrating relationship between the processed food and services prices is also found by Espasa et al. (2002). The (un)processed food prices as explanatory endogenous variables for the models of processed food and services, respectively, are forecast according to the corresponding model specifications, even though the optimal models for the food prices seem to be different. Using different models for food prices might reduce positive correlation among forecast errors of these three components of HICP.

For both areas, the small number of lags selected for all models is noticeable. In previous specifications lag lengths of up to 12 were included, but it seems that the few significant coefficients with longer lags are not very stable. Over the latest sample, especially the selection criteria with relatively strong penalty for extra parameters suggested at most one lag. For services and the total index, the 12th lag is significant as well.

11 The German import price index is used as no monthly Dutch import price index is available.

TABLE 1 – NETHERLANDS: HICP (SUB)INDICES

HICP-index	P^{uf}	P^{pf}	P^{i}	P^{e}	P^{s}	P^{total}
Exogenous	–	Wages^{NL}, $e^{€\$}$	Wages^{NL}, P^{oil}	P^{oil}	Wages^{NL}, $e^{€\$}$	Wages^{NL}, $e^{€\$}$, P^{oil}
Endogenous	–	Pim^{GE}	Pim^{GE}	–	Pim^{GE}	Pim^{GE}
EC-term	–	P^{pf} Wages^{NL} Pim^{GE}	P^{i} Wages^{NL} Pim^{GE}	–	P^{s} Wages^{NL}, Pim^{GE}	P^{total} Wages^{NL} Pim^{GE}
Lags included	0	1	0	0	1, 12	12
Specification*	Δ_1	$\Delta_1\Delta_{12}$	$\Delta_1\Delta_{12}$	Δ_1	$\Delta_1\Delta_{12}$	$\Delta_1\Delta_{12}$

*Δ_x is defined as the x month difference of the variable. The error-correction (EC-)term is specified in annual inflation rates. The models in first differences include seasonal dummies

TABLE 2 – EURO AREA: HICP (SUB)INDICES

HICP-index	P^{uf}	P^{pf}	P^{i}	P^{e}	P^{s}	P^{total}
Exogenous	–	Wages^{EU} $e^{€\$}$	Wages^{EU}, $e^{€\$}$	P^{oil}	Wages^{EU}	Wages^{EU}, $e^{€\$}$, p^{oil}, wmp^{exe}
Endogenous	–	P^{uf}	P^{prod}	P^{prod}	P^{pf}	P^{prod},
EC-term	–	–	P^{i} P^{prod}, Wages^{EU}	–	P^{s} P^{pf}, Wages^{EU}	P^{total}, P^{prod}, Wages^{EU}
Lags included	1	1	0	1	1, 12	12
Specification*	Δ_1	Δ_1	$\Delta_1\Delta_{12}$	Δ_1	$\Delta_1\Delta_{12}$	$\Delta_1\Delta_{12}$

*Δ_x is defined as the x month difference of the variable. The error-correction (EC-)term is specified in annual inflation rates. The models in first differences include seasonal dummies

4 FORECAST UNCERTAINTY

The constructed models provide a conditional forecasts for the inflation rates in the short to medium-term. We will supplement the point forecasts generated by the models with prediction intervals that provide a quantitative content for the uncertainty surrounding them. The Bank of England quantifies uncertainty by publishing[12] density forecasts, which is an estimate of the complete probability distribution of the possible future values of a variable, see also Wallis (1999). In this study, we will use non-parametric bootstrapping to construct a probability distribution and deduce the corresponding prediction interval, see Horowitz (2001). We perform a simulation experiment in which the error terms are drawn from the distribution of the residuals of the

12 The Bank of England has published a density forecast of inflation in its quarterly *Inflation Report* since February 1996.

estimated models. The simulation draws from the multi-variate empirical distribution to preserve the contemporaneous interdependence of the residuals of the five models. If the bootstrap procedure would be computed for all five categories separately, the overall HICP confidence band would become narrower due to the neglected positive correlation among components.[13] Wages, the exchange rate and the oil and commodity prices are exogenous explanatory variables in our models by assumption. The uncertainty surrounding these exogenous predictions is not taken into account.

The bootstrapped distribution of the inflation forecasts turns out to be fairly symmetric, see Figure 1. The graph shows the forecasts for the Netherlands and the euro area over 2002 and the first half of 2003, based on the models given in Table 1, respectively, 2 and the realisations for the exogenous variables. For both areas, two forecasts and corresponding confidence bands are given: one based on aggregation of the models for the components, P^{agg}[14], and one based on the model for total HICP inflation, P^{total}. Both model forecasts are in the centre of the confidence band. The bootstrap median and mode are almost identical to its mean implying an almost symmetrical bootstrap distribution. In the past, for the Netherlands we sometimes found that the point forecast was not in the middle of the confidence band. The difference between the two was probably due to a bias in the AR-coefficients. The bias correction methodology of Kilian (1998), which implies using the bootstrap twice, might reduce the difference between the two under those circumstances.

Although the models are estimated with data up until 2001, this exercise is not fully out-of-sample in the sense that the data for 2002 was previously included to select the optimal models. Despite these facts, the realised inflation in January 2002 for the euro area was clearly above the prediction interval for the total HICP model and just on the border for the aggregate model. The most likely reason for this under-estimation of inflation seems to be the cash changeover to the euro. This event is not explicitly incorporated in these forecasts.

An attractive feature of both models for both areas is that the forecast accuracy does not deteriorate over the forecast horizon. Probably, the inclusion of the error correction term including exogenous variables guarantees reasonable long-term forecasts. Of course, this is only true in as far as we are able to predict the exogenous variables with reasonable accuracy.

With respect to the difference between the aggregated and the total HICP model results, for both areas the aggregated results seem better. First of all, they

13 Moreover, the confidence band is widened further since in the bootstrap procedure adds an additional residual to the prediction in order to reflect the future uncertainty of unforeseeable shocks.

14 The weights in the aggregation are updated yearly. Over the forecast horizon, they are assumed constant.

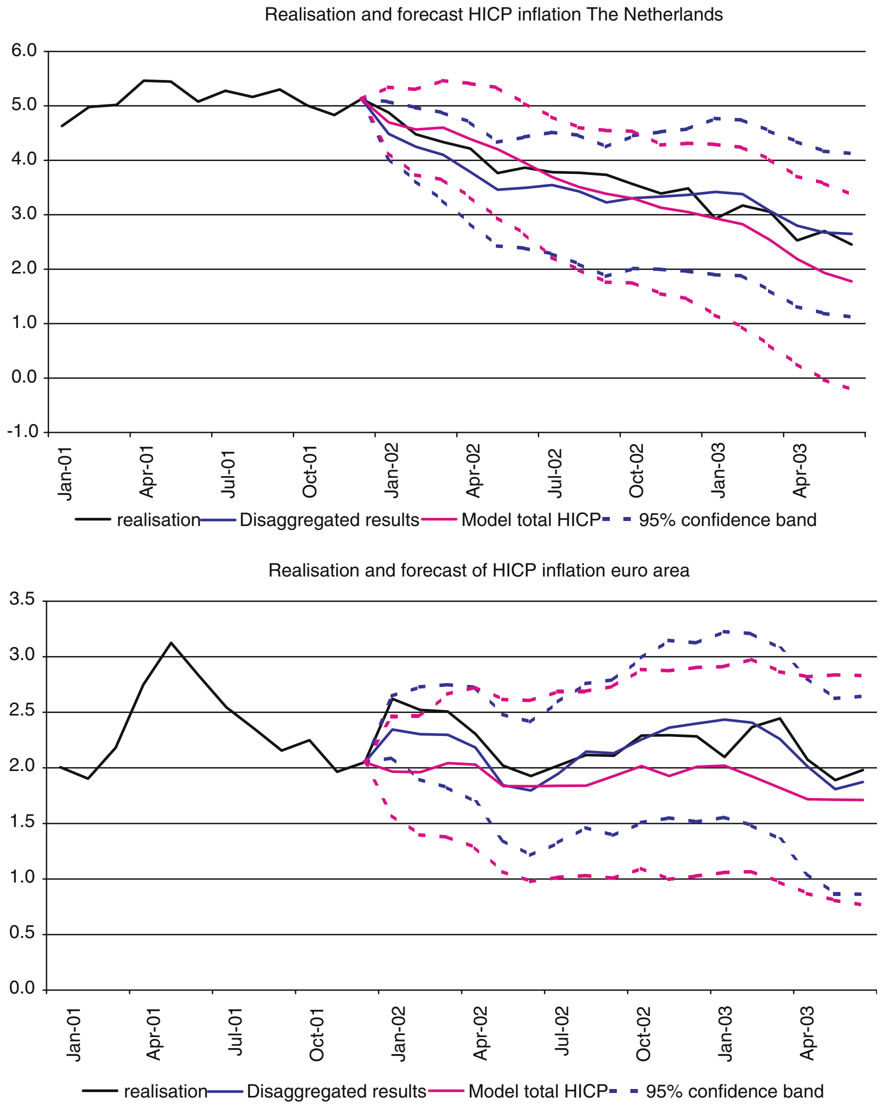

Figure 1 – Realisation and forecast of HICP inflation for the Netherlands and the euro area

follow the realised inflation more closely. Moreover, the confidence bands for the total HICP models are wider, especially for longer forecast horizons. However, this forecast evaluation is only based on one period. In the next section, we will evaluate the forecast performance of all models more systematically.

5 MODEL EVALUATION

In order to systematically evaluate the models, we have computed the root mean squared error of recursive dynamic out-of-sample forecasts (Stock and Watson (1999)), for the model specifications given in Tables 1 and 2. For this purpose, the realisations of the exogenous variables are used. For the Netherlands this includes gas prices, housing rents and radio and television (RTV) licences. The latter is taken as exogenous as the abolition of them in January 2000 had a huge negative impact on services inflation in that year (see Figure A1). If no account is taken of this event, the fit of the models deteriorate. Our first out-of-sample evaluation is for the period 1998:1 up until 1999:6 based on data up until 1997:12, whereas the last exercise involves 2003:1 up until 2004:6, leading to 61 recursive forecast errors for each horizon.

Tables 3 and 4 show the RMSFE for the Netherlands, respectively, the euro area. The forecast errors are evaluated for the year-on-year percentage change in the respective HICP component. The forecast errors of the estimated models are compared to those of a naive forecast, which sets all the forecasts ahead equal to the latest observed annual inflation rate, and optimal univariate AR models. The AR models are in first differences, with seasonal dummies, and the lag length is chosen based on the Schwarz criterion for the full sample. With respect to the total HICP, both the results of the own model for total HICP and the one based on an aggregate of components are reported.

For both areas, we find that the models outperform the naive forecast almost uniformly. For the Netherlands, this is always the case, whereas for the euro area the up to three periods ahead model forecast for P^i and the up to 12 periods ahead model forecasts for P^{pf} are about equally good as the naive ones. Although, outperforming the naive forecast seems hardly demanding, the results for the optimal AR models for P^i and P^s for the euro area and the ones for P^{total} and P^{agg} show it is far from trivial. Relative to both benchmarks, the models perform very good, although the models for P^{pf}, and P^i are slightly outperformed by the AR models for short horizons.

Comparing the naive forecast errors for the Netherlands with those for the euro area, it is clear that Dutch inflation is much more volatile than inflation in the total euro zone. Many of the shocks to inflation are country-specific and these shocks partly cancel for the euro area. The exception is energy inflation. For energy, oil prices are most important and these shocks hit all countries at the same time. For the models, the difference in RMSFE between the two areas is less extreme. Consequently, the improvement relative to the naive forecast is bigger for the models for the Netherlands than for those of the euro area. For forecasts seven or more months ahead, the Dutch model RMSFE for energy is even smaller than the corresponding euro area one. This is probably due to the assumption of exogenous natural gas prices.

TABLE 3 – NETHERLANDS: RECURSIVE ROOT MEAN SQUARED FORECAST ERROR 1998–2002, 1–18 MONTHS AHEAD

	P^{uf}			P^{pf}			P^{i}			P^{e}			P^{s}			P^{total}			P^{agg}	
Horizon	Naive	AR 0	Model	Naive	AR (1)	Model	Naive	AR (12)	Model	Naive	AR (0)	Model	Naive	AR (0)	Model	Naive	AR (0)	Model	AR	Model
1	*1.93*	**1.48**	**1.48**	*0.42*	0.30	0.39	*0.44*	**0.40**	0.44	*1.92*	1.62	**0.92**	*0.38*	0.35	**0.23**	0.36	*0.37*	0.31	0.35	**0.28**
2	*2.75*	**2.19**	**2.19**	*0.68*	**0.49**	0.62	*0.61*	**0.55**	0.59	*2.75*	2.35	**1.29**	*0.56*	0.52	**0.33**	0.51	*0.56*	0.41	0.53	**0.39**
3	*3.37*	**2.71**	**2.71**	*0.89*	**0.66**	0.76	*0.71*	**0.64**	0.69	*3.27*	**2.91**	**1.49**	*0.68*	0.67	**0.40**	0.62	*0.71*	0.50	0.67	**0.45**
4	*3.88*	**3.12**	**3.12**	*1.07*	**0.84**	0.87	*0.84*	**0.74**	0.80	*3.87*	3.35	**1.70**	*0.80*	0.78	**0.46**	0.76	*0.85*	0.60	0.81	**0.55**
5	*4.27*	**3.47**	**3.47**	*1.23*	0.99	**0.96**	*1.01*	**0.87**	0.90	*4.70*	3.92	**1.99**	*0.90*	0.89	**0.50**	0.90	*0.98*	0.67	0.93	**0.63**
6	*4.71*	**3.77**	**3.77**	*1.38*	1.15	**1.07**	*1.18*	**0.99**	0.99	*5.29*	4.36	**2.09**	*0.99*	0.97	**0.51**	1.00	*1.09*	**0.72**	1.04	**0.69**
7	*5.10*	**4.02**	**4.02**	*1.53*	1.31	**1.16**	*1.32*	1.13	**1.05**	*5.88*	4.77	**2.12**	*1.10*	1.04	**0.50**	1.10	*1.19*	0.75	1.12	**0.73**
8	*5.47*	**4.21**	**4.21**	*1.66*	1.45	**1.20**	*1.46*	1.27	**1.09**	*6.48*	5.16	**2.25**	*1.20*	1.10	**0.50**	1.21	*1.29*	0.76	1.21	0.77
9	*5.90*	**4.49**	**4.49**	*1.80*	1.58	**1.23**	*1.57*	1.37	**1.13**	*7.01*	5.58	**2.40**	*1.30*	1.16	**0.49**	1.32	*1.39*	**0.76**	1.30	0.81
10	*6.36*	**4.73**	**4.73**	*1.92*	1.71	**1.26**	*1.68*	1.47	**1.16**	*7.36*	5.84	**2.39**	*1.41*	1.23	**0.49**	1.42	*1.48*	**0.74**	1.37	0.85
11	*6.86*	**4.91**	**4.91**	*2.06*	1.85	**1.29**	*1.80*	1.57	**1.19**	*7.77*	6.12	**2.40**	*1.51*	1.31	**0.49**	1.52	*1.57*	**0.73**	1.45	0.89
12	*7.31*	**5.14**	**5.14**	*2.20*	1.99	**1.32**	*1.93*	1.69	**1.21**	*8.22*	6.43	**2.46**	*1.60*	1.39	**0.52**	1.63	*1.68*	**0.72**	1.55	0.95
13	*7.56*	**5.11**	**5.11**	*2.31*	2.05	**1.31**	*2.03*	1.74	**1.23**	*8.50*	6.39	**2.43**	*1.66*	1.39	**0.56**	*1.71*	1.69	**0.71**	1.57	0.98
14	*7.68*	5.07	5.07	*2.43*	2.11	1.31	*2.11*	1.78	**1.23**	*8.77*	6.40	**2.40**	*1.73*	1.40	**0.63**	*1.79*	1.70	**0.74**	1.60	1.01
15	*7.79*	**5.05**	**5.05**	*2.53*	2.13	**1.30**	*2.21*	1.83	**1.24**	*9.09*	6.40	**2.37**	*1.79*	1.40	**0.69**	*1.87*	1.72	**0.77**	1.62	1.05
16	*7.83*	**5.05**	**5.05**	*2.62*	2.14	**1.30**	*2.30*	1.86	**1.25**	*9.39*	6.42	**2.35**	*1.84*	1.41	**0.75**	*1.94*	1.73	**0.82**	1.64	1.09
17	*7.81*	**5.05**	**5.05**	*2.71*	2.16	**1.29**	*2.37*	1.90	**1.26**	*9.57*	6.45	**2.35**	*1.89*	1.41	**0.80**	*2.00*	1.73	**0.88**	1.66	1.12
18	*7.81*	**5.07**	**5.07**	*2.81*	2.18	**1.28**	*2.44*	1.93	**1.25**	*9.73*	6.47	**2.35**	*1.92*	1.41	**0.87**	*2.06*	1.74	**0.96**	1.67	1.15

The forecast errors are computed over the annual inflation rates. The models for services and total HICP are corrected for the abolition of RTV licences. The lag length of the AR models is based on the Schwarz criterion using the full sample. The lowest RMSFE for each index is printed in bold face, the highest one in italics

TABLE 4 – EURO AREA: RECURSIVE ROOT MEAN SQUARED FORECAST ERROR 1998–2002, 1–18 MONTHS AHEAD

	P^{uf}			P^{pf}			P^{i}			P^{e}			P^{s}			P^{total}			P^{agg}	
Horizon	Naive	AR (1)	Model	Naive	AR (3)	Model	Naive	AR (12)	Model	Naive	AR (0)	Model	Naive	AR (12)	Model	Naive	AR (12)	Model	AR	Model
1	*0,80*	**0,60**	**0,60**	*0,16*	**0,13**	0,14	**0,21**	*0,24*	0,21	*1,87*	1,40	**0,87**	0,19	*0,22*	**0,17**	*0,21*	0,20	0,18	0,20	**0,15**
2	*1,36*	**0,99**	**0,99**	*0,25*	**0,21**	0,23	**0,33**	*0,36*	0,33	*2,79*	2,09	**1,07**	0,21	*0,26*	**0,19**	*0,32*	0,29	0,27	0,30	**0,23**
3	*1,78*	**1,30**	**1,30**	*0,36*	**0,28**	0,30	**0,38**	*0,40*	0,39	*3,56*	2,65	**1,32**	0,26	*0,31*	**0,22**	*0,39*	0,32	0,33	0,36	**0,27**
4	*2,14*	**1,57**	**1,57**	*0,46*	**0,37**	0,38	0,40	*0,42*	**0,40**	*4,39*	3,30	**1,59**	0,29	*0,33*	**0,24**	*0,43*	0,33	0,36	0,41	**0,29**
5	*2,42*	**1,80**	**1,80**	*0,54*	**0,44**	0,46	0,40	*0,42*	**0,39**	*5,05*	3,73	**1,79**	0,34	*0,38*	**0,28**	0,43	0,33	0,38	*0,44*	**0,32**
6	*2,61*	**1,97**	**1,97**	*0,62*	**0,51**	0,54	0,40	*0,42*	**0,38**	*5,68*	4,13	**2,00**	0,39	*0,45*	**0,31**	0,43	0,35	0,40	*0,47*	**0,34**
7	*2,81*	**2,14**	**2,14**	*0,69*	**0,59**	0,62	0,45	*0,51*	**0,41**	*6,31*	4,56	**2,23**	0,45	*0,51*	**0,34**	0,44	0,40	0,42	*0,51*	**0,37**
8	*3,04*	**2,29**	**2,29**	*0,76*	**0,68**	0,69	0,51	*0,61*	**0,46**	*6,89*	4,90	**2,41**	0,49	*0,56*	**0,36**	0,45	0,45	0,45	*0,57*	**0,41**
9	*3,33*	**2,49**	**2,49**	*0,82*	**0,76**	0,77	0,55	*0,65*	**0,48**	*7,48*	5,23	**2,61**	0,54	*0,62*	**0,39**	0,48	0,50	0,49	*0,61*	**0,45**
10	*3,65*	**2,71**	**2,71**	*0,89*	**0,85**	0,85	0,57	*0,67*	**0,48**	*8,07*	5,59	**2,80**	0,58	*0,66*	**0,42**	0,52	0,56	0,55	*0,67*	**0,50**
11	*3,97*	**2,91**	**2,91**	*0,95*	**0,94**	0,94	0,60	*0,67*	**0,47**	*8,61*	5,93	**2,92**	0,62	*0,70*	**0,46**	0,57	0,61	0,60	*0,72*	**0,54**
12	*4,25*	**3,10**	**3,10**	**1,02**	1,02	*1,03*	0,62	*0,65*	**0,46**	*9,20*	6,33	**3,06**	0,67	*0,74*	**0,49**	0,63	0,68	0,65	*0,77*	**0,58**
13	*4,43*	**3,13**	**3,13**	1,07	*1,07*	**1,06**	0,64	*0,68*	**0,46**	*9,61*	6,29	**3,11**	0,70	*0,78*	**0,54**	0,65	0,71	0,69	*0,77*	**0,59**
14	*4,57*	**3,13**	**3,13**	*1,10*	1,10	**1,08**	0,68	*0,72*	**0,46**	*9,90*	6,25	**3,12**	0,74	*0,83*	**0,58**	0,67	0,73	0,73	*0,77*	**0,59**
15	*4,66*	**3,14**	**3,14**	*1,14*	**1,11**	**1,08**	0,70	*0,73*	**0,45**	*10,21*	6,23	**3,11**	0,77	*0,86*	**0,61**	0,68	0,76	0,77	*0,77*	**0,60**
16	*4,71*	**3,14**	**3,14**	*1,18*	1,14	**1,11**	0,70	*0,74*	**0,42**	*10,42*	6,23	**3,11**	0,81	*0,91*	**0,65**	0,69	*0,79*	*0,81*	0,77	**0,60**
17	*4,74*	**3,14**	**3,14**	*1,21*	1,16	**1,13**	0,70	*0,74*	**0,39**	*10,57*	6,22	**3,09**	0,83	*0,95*	**0,67**	0,71	*0,83*	*0,86*	0,78	**0,61**
18	*4,78*	**3,14**	**3,14**	*1,24*	1,17	**1,15**	0,71	*0,75*	**0,37**	*10,72*	6,23	**3,05**	0,86	*0,99*	**0,71**	0,73	*0,88*	*0,91*	0,79	**0,61**

The forecast errors are computed over the annual inflation rates. The lag length of the AR models is based on the Schwarz criterion using the full sample. The lowest RMSFE for each index is printed in bold face, the highest one in italics

With respect to the advantage of splitting up the HICP index to forecast total HICP inflation, the results are somewhat mixed. For the Netherlands, we find that aggregation of components leads to a lower RMSFE for forecasts up to 7 months ahead, whereas for longer forecast horizons the opposite holds. For the euro area, the aggregation method performs best for all forecast horizons. These results imply that the dominance of the indirect approach for 2002 and the first half of 2003, which appeared clear from Figure 1, seems to be a general feature. The relative good performance of the aggregation method for all forecast horizons runs counter to the results of Fritzer et al. (2002). For VAR models, they found the direct approach to perform better for horizons up to 9 months ahead, after which the aggregation approach was to be preferred. Hubrich (2001, 2005) and Benalal et al. (2004) on the other hand found that aggregation performed especially worse at long horizons. Also with respect to the AR models, no common feature is found. Whereas for the Netherlands the disaggregated approach produces better results, the opposite holds for the euro area. In general, it seems that forecast errors among HICP sub-indices are too positively correlated to be able to gain a lot by aggregating component models.

The relatively good forecast performance of our models does of course depend on our ability to predict exogenous variables correctly. In Tables 3 and 4, the realisations are used to make forecasts, but obviously these are not available when making really out-of-sample forecasts. For the Netherlands, we do have a way to check the relevance of this objection as the Dutch models have been used for the NIPE since December 1998. Consequently, we have totally out-of-sample forecasts for HICP inflation and its five components for 16 forecasting rounds.

In Figure 2, the root mean squared forecast errors of the NIPE projections are shown together with the ones for the naive forecast, the optimal AR models and those generated with the currently selected models using realisations for the exogenous variables. The squared forecast errors are averaged over all projections that were made for a certain forecast horizon, that is 16 for 1–11 months ahead, 12 for 12 months ahead, 8 for 13 and 14 months ahead and 3 for 15 months ahead. This explains the sudden drop in RMSFE at horizon 15.

For every HICP component, the NIPE projections outperformed the naive and AR benchmarks almost uniformly. Compared to the model cum realised exogenous variables forecasts the results are more mixed. In principle, there are three reasons for differences between the model forecasts and the NIPE results. First, the assumptions regarding the exogenous variables differ. Second, different models were used. Third, the NIPE also includes 'add-factors' to account for judgmental issues. Unfortunately, we do not have a complete track record of the models and assumptions used. Otherwise, we could identify the exact relevance of each of the three factors. Nevertheless, some

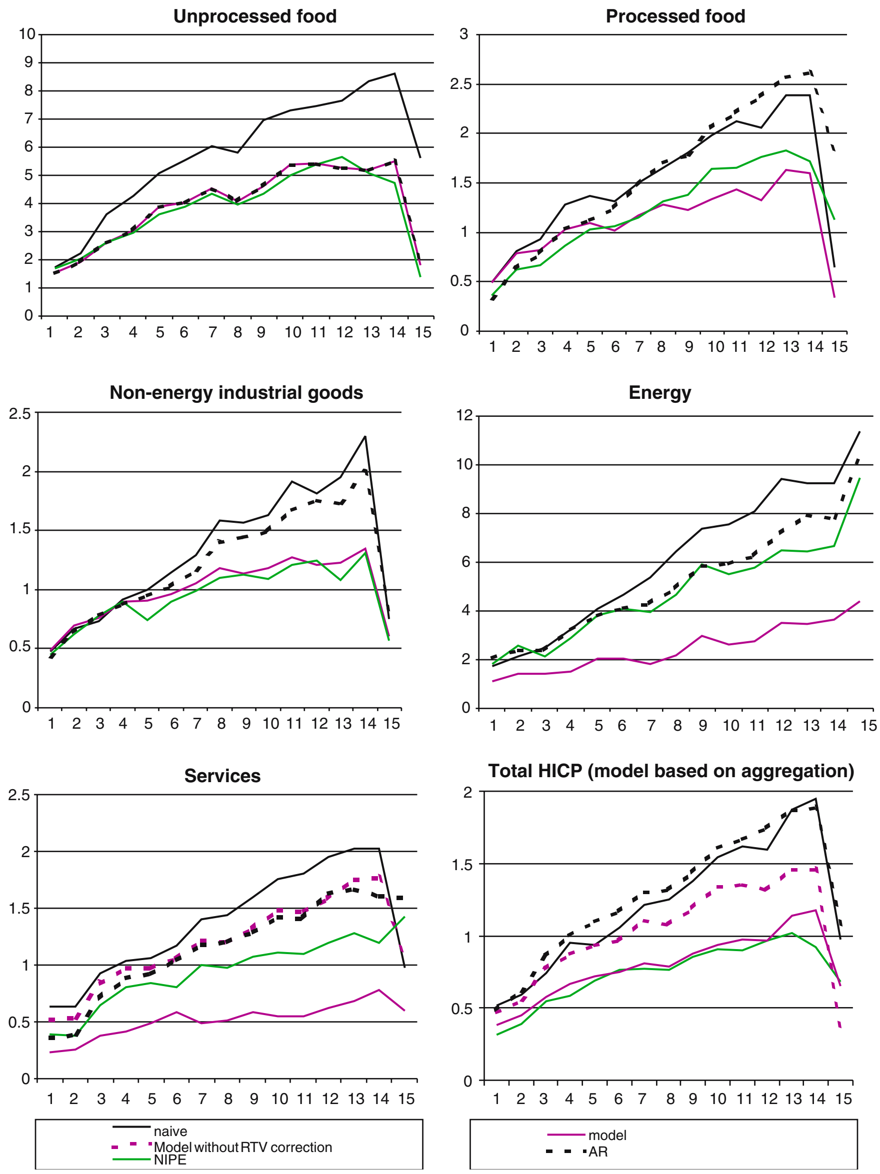

Figure 2 – Root mean squared forecast error Dutch HICP inflation 1–15 months ahead

interesting conclusions can be drawn. Of the three factors, wrong assumptions probably only lead to worse predictions. For the model specification it can go either way, whereas judgement hopefully only improves the results.

The only two sub-indices for which the NIPE performed systematically worse than the model are energy and services. For energy, this is not at all surprising as the development of oil prices is highly unpredictable and very influential on P^e. Moreover, gas prices, which account to almost 40% of Dutch energy budget, also sometimes moved more than expected. With respect to services, the result is partly due to the abolition of RTV licences in January 2000, which was not foreseen in the NIPE projections of 1999. The figure also shows the impact of ignoring this event for our selected model. It is clear that the model falls apart, as the NIPE now outperforms the model. Besides this effect, the housing rents, which accounted for about 30% of the Dutch services budget, were not always perfectly predictable.

Given the relatively big forecast errors for energy and services, it is surprising to see that for the overall index P^{agg} the NIPE performs even slightly better than the model forecast, even though both are based on the sub-indices. Apparently, the correlation among HICP components was higher for models with correct exogenous variables than for the NIPE. Consequently, the disadvantage of not knowing the future values of exogenous variables was compensated enough by the possibility to add judgement.

6 CONCLUSION

This paper describes the procedures we use to predict monthly Dutch and euro area HICP inflation. The HICP prediction is constructed by aggregating forecasts for the five HICP sub-indices unprocessed food, processed food, non-energy industrial goods, energy and services, whereas total HICP is also modelled directly for comparison reasons. All models are linear vector autoregressive or error correction models, possibly including exogenous variables.

In order to select the appropriate models, the first step is a visual inspection of the data. Those price indices, which show a clear changing seasonal pattern are modelled in both first and 12 month differences including an error correction term representing long-run equilibrium relationships between inflation and other variables (if they have the correct sign and reasonable order of magnitude). Price indices without clear structural breaks in seasonal pattern (unprocessed food and energy) are modelled in first differences. Here, no error correction term is included. The second step involves the calculation of all potential models, using a small set of exogenous and endogenous variables. We select the best models according to nine different statistical selection criteria, using both in-sample goodness-of-fit, parsimony and out-of-sample forecasting accuracy. In the third step, the optimal models are chosen, based both on the statistical criteria and economic evaluation. Especially, the long-run properties are important here. Expected wage developments form a very important anchor in this respect.

Once an appropriate model is chosen, all available data are used to generate forecasts. Foreseeable shocks over the forecast horizon (for instance a change in indirect taxes) are incorporated ex-post. According to a recursive root mean squared forecast error evaluation exercise, all models outperform the naive forecast and optimal AR models on most forecast horizons. The comparison between the errors of forecasting the aggregate directly or aggregating the forecasts of the components shows a clear preference for aggregating in the euro area. For the Netherlands for short forecast horizons aggregating is better, but for longer horizons the direct approach is to be preferred. These evaluations do depend on perfect knowledge of future values for exogenous variables however.

For the Netherlands, a fully out-of-sample exercise is performed by evaluating the first 16 NIPE rounds. The forecast performance of the NIPE projections is even slightly better than the one for the selected models with perfect foresight of exogenous variables. Again, the naive forecast is outperformed on every forecast horizon for every (sub-)index. Apparently, judgement more than compensates for the lack of knowledge on the future values of exogenous variables. Indeed, forecasting inflation seems to be an art as well as a science!

Overall, the robustness of the inflation forecasting models, both with respect to the selection criterion used and over time, is not encouraging though. The optimal model is not likely to exist, making regular evaluation of models and the permanent good use of common sense all the more important.

APPENDIX DATA

The sample period of the data set is October 1987, respectively, January 1990 for the Netherlands and the euro area until August 2002. Table A1 lists all the variables that are currently included. Apart from these selected variables, other variables have been tested but are not selected in the final models.

TABLE A1 – DATA, NOTATION AND SOURCE CODE

Variable	Notation	External source
Harmonised index of consumption prices		
Euro area		
HICP	Pea^{total}	Eurostat
HICP unprocessed food	Pea^{uf}	Eurostat
HICP procecessed food	Pea^{pf}	Eurostat
HICP industrial production excl. Energy	Pea^{i}	Eurostat
HICP energy	Pea^{e}	Eurostat
HICP services	Pea^{s}	Eurostat
Netherland		
HICP	Pnl^{total}	Eurostat

TABLE A1 – continued

HICP unprocessed food	Pnl^{uf}	Eurostat
HICP processed food	Pnl^{pf}	Eurostat
HICP industrial goods excl. Energy	Pnl^{i}	Eurostat
Variable	Notation	External source
HICP energy	Pnl^{e}	Eurostat
HICP services	Pnl^{s}	Eurostat
Endogenous variables		
Import price index Germany	Pim^{GE}	Federal statistical office germany
Producer prices (euro area)	P^{prod}	BIS
Exogenous variables		
Euro/dollar exchange rate	$e^{€\$}$	ECB
Oil price (Brent crude) in euro	P^{oil}	IFS / Bloomberg [a)]
Hourly wages industry, euro area	$\mathrm{Wages}^{\mathrm{EU}}$	[b)]
Hourly wages private sector, Netherlands	$\mathrm{Wages}^{\mathrm{NL}}$	CBS
Commodity prices (excl.energy) in euro	$\mathrm{Wmp}^{\mathrm{exe}}$	HWWA [a)]

[a)] Recent data as well as projections for the forecast horizon are obtained from the ECB. The projections are based on futures prices

[b)] The euro area hourly wage is an average of the individual country's hourly wage rates, weighted by the GDP-share in 1995. For Belgium, Denmark, Spain, Finland, France, Greece and Ireland only data on quarterly basis is available. This is interpolated to monthly data by the Lisman-procedure. Portugal and Luxembourg are not considered due to lack of data. Moreover, a 12-months centered moving average is applied to smooth the aggregated hourly wage rate in order to get more reliable parameter estimates

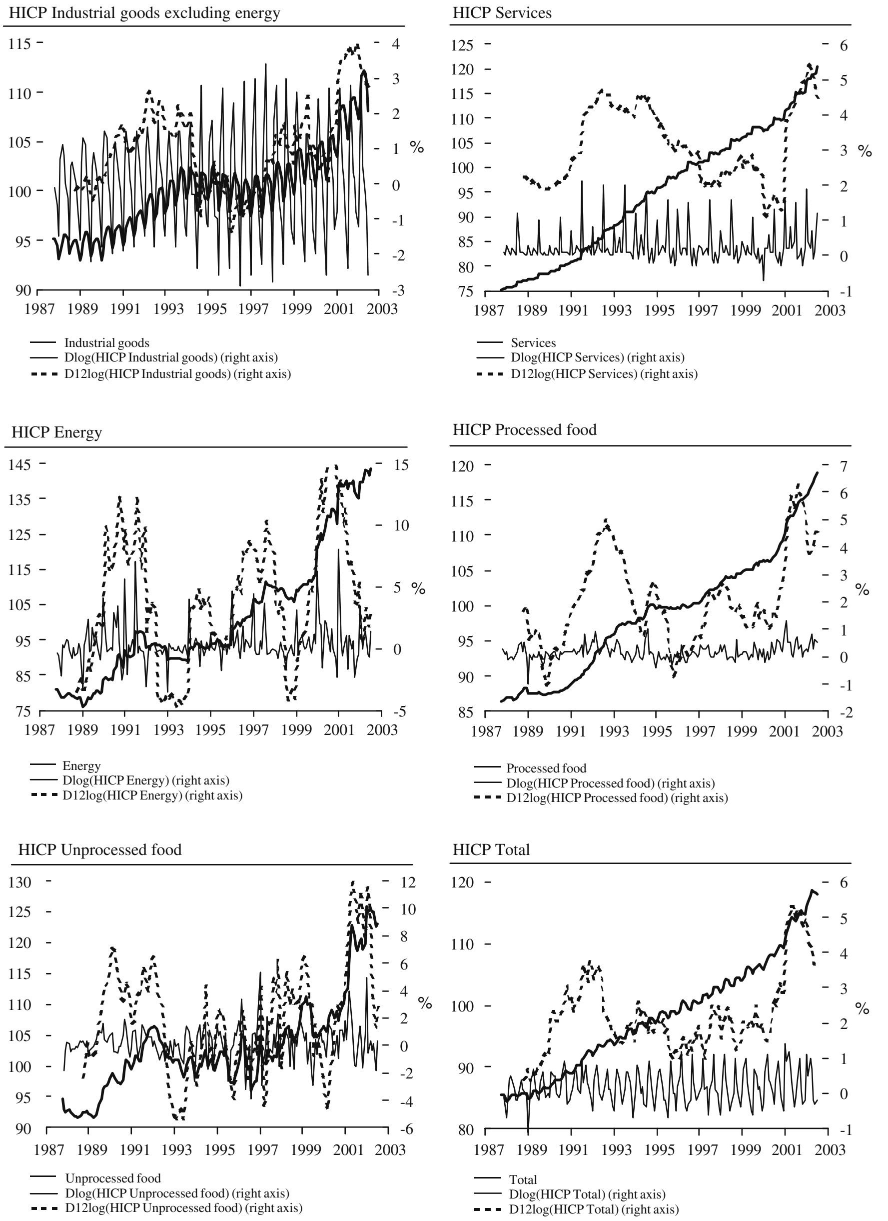

Figure A1 – HICP (sub)indices in original, monthly and annual inflation format for the Netherlands

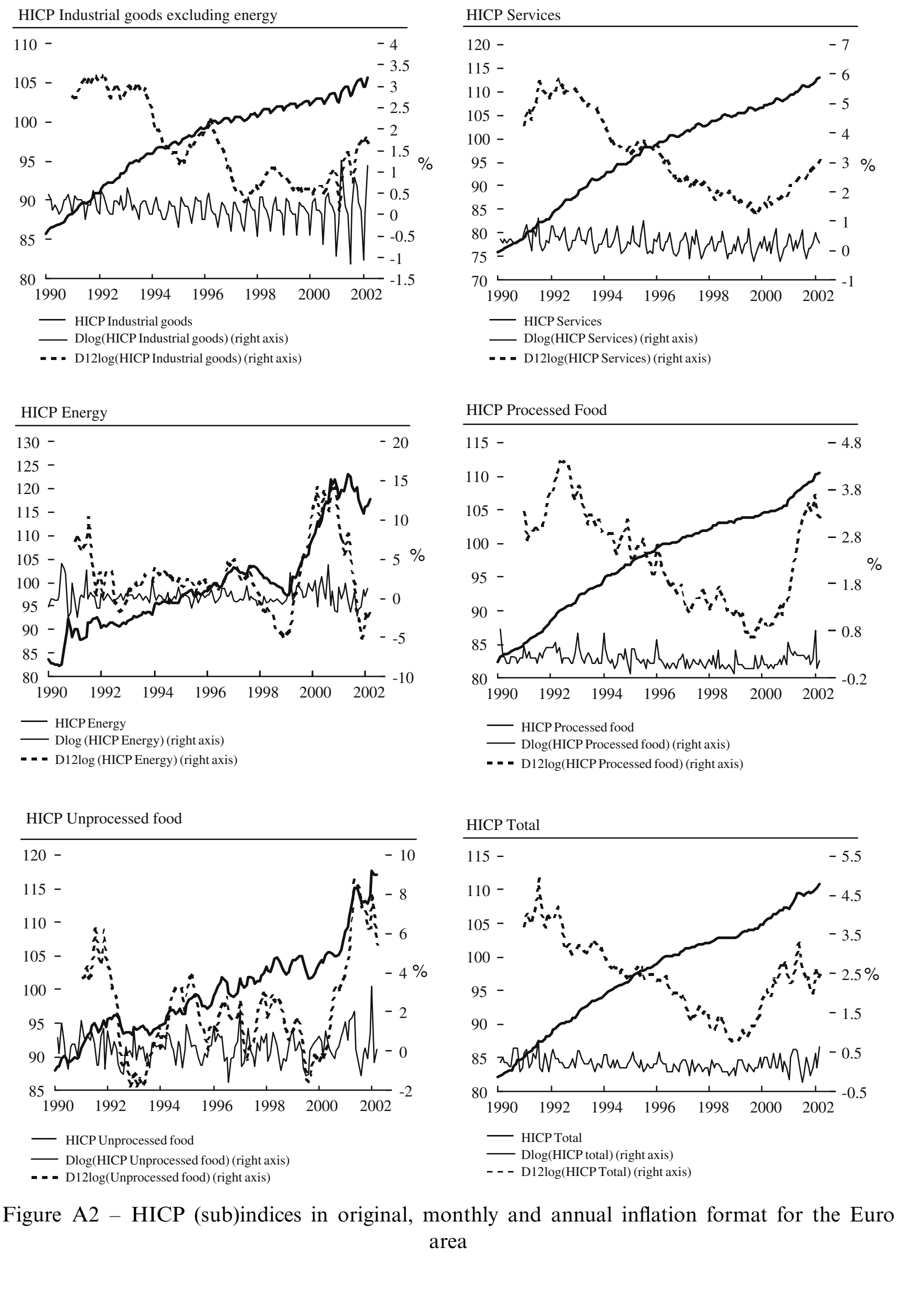

Figure A2 – HICP (sub)indices in original, monthly and annual inflation format for the Euro area

REFERENCES

Benalal, N., J.L. Diaz del Hoyo, B. Landau, M. Roma and F. Skudelny (2004), 'To Aggregate or not to Aggregate? Euro Area Inflation Forecasting,' ECB Working Paper, 374.

Banerjee, A., M. Marcellino and I. Masten (2003), 'Leading Indicators for Euro Area Inflation and GDP Growth,' CEPR Discussion Paper, 3893.

Bruneau, C., O. de Bandt and A. Flageollet (2003a), 'Forecasting Inflation in the Euro Area,' Working Paper Banque de Franc, 102.

Bruneau, C., O. de Bandt, A. Flageollet and E. Michaux (2003b), 'Forecasting Inflation Using Economic Indicators: The Case of France,' Working Paper Banque de France, 101.

Christoffersen, P.F. and F.X. Diebold (1998), 'Cointegration and Long-Horizon Forecasting,' *Journal of Business and Economic Statistics*, 16, p. 450–458.

Espasa, A., E. Senra and R. Albacete (2002), 'Forecasting Inflation in the European Monetary Union: A Disaggregated Approach by Countries and by Sectors,' *The European Journal of Finance*, 8, p. 402–421.

Fritzer, F., G. Moser and J. Scharler (2002), 'Forecasting Austrian HICP and its Components Using VAR and ARIMA models,' Working Paper OENB, 73.

Hendry, D.F. and G.E. Mizon (1999), 'On Selecting Policy Analysis Models by Forecast Accuracy,' University of Southampton Discussion Papers in Economics and Econometrics, 9918.

Horowitz, J.L. (2001), 'The Bootstrap, in: J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics*,' Vol. 5, Elsevier Science B.V., Amsterdam, Ch. 52, p. 3159–3228.

Hubrich, K. (2001), Forecasting Euro Area Inflation: Does Contemporaneous Aggregation Improve the Forecasting Performance?,' Research Memorandum WO&E, 661.

Hubrich, K. (2005), 'Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast accuracy?,' *International Journal of Forecasting*, 21(1), p. 119–136.

Inoue, A. and L. Kilian (2006), 'On the Selection of Forecasting Models,' *Journal of Econometrics*, Forthcoming.

Kilian, L. (1998), 'Small-Sample Confidence Intervals for Impulse Response Functions,' *Review of Economics and Statistics*, 80(2), p. 218–230.

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

Marcellino, M., J.H. Stock and M.W. Watson (2003), 'Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-Wide Information,' *European Economic Review*, 47(1), p. 1–18

Moser, G., F. Rumler and J. Scharler (2004), 'Forecasting Austrian Inflation,' Working Paper OENB, 91.

Stock, J.H. and M.W. Watson (1999), 'Forecasting Inflation,' *Journal of Monetary Economics*, 44(2), p. 293–335.

Stock, J.H. and M.W. Watson (2003), 'Forecasting Output and Inflation: The Role of Asset Prices,' *Journal of Economic Literature*, XLI, p. 788–829.

Wallis, K.F. (1999), 'Asymmetric Density Forecasts of Inflation and the Bank of England's Fan Chart,' *National Institute Economic Review*, 167, p. 106–112.