Royal Institute of Technology

Master Thesis

Forecasting Euro Area Inflation
By Aggregating Sub-components

Author:
Noah Clason

Supervisors:
Peter Kaplan
Tatjana Pavlenko

Department of Mathematical Statistics

February 24, 2013

Abstract

The aim of this paper is to see whether one can improve on the naive forecast of Euro Area inflation, where by naive forecast we mean the year-over-year inflation rate one-year ahead will be the same as the past year. Various model selection procedures are employed on an autoregressive-moving-average model and several Phillips curve based models. We test also if we can improve on the Euro Area inflation forecast by first forecasting the sub-components and aggregating them. We manage to substantially improve on the forecast by using a Phillips curve based model. We also find further improvement by forecasting the sub-components first and aggregating them to Euro Area inflation.

Contents

1 Introduction

The European Central Bank (ECB) primary objective is to maintain price stability for the Euro Area, which they define as "a year-on-year increase in the Harmonized Index of Consumer Prices (HICP) for the Euro Area of below 2%". They also have clarified the definition to "in the pursuit of price stability, it aims to maintain inflation rates below, but close to, 2% over the medium term." Therefore we focus on forecasting Euro Area HICP year-over-year (Y-o-Y) rate one-year ahead.

The Phillips curve is the mainstream inflation forecasting model and offers the best framework for understanding monetary policy. However, it's well known that forecasting inflation is notoriously hard; Atkeson and Ohanian (2001) showed that the naive forecast, that inflation over the next year will be the same as it has been during the past year, performed better or just as well as the three standard Phillips curve-based models they examined.

This paper focuses on the different methodological decisions and issues one faces when wanting to perform an inflation forecast using a univariate or Phillips curve based model. For example, should one use lags of Y-o-Y changes instead of M-o-M (month-over-month) changes? Should one strictly follow the Phillips curve and model inflation rate as a unit-root process? Does an iterated or direct forecast perform better? To answer these questions we perform out-of-sample forecasts trying to simulate the real-time forecasting experience, using the data available at the time for both model selection and estimation.

We then use our set of models to see whether forecasting the sub-components of HICP and then aggregating up to Euro Area HICP produces better forecasts than to directly forecast Euro Area HICP. Several researchers have done this exercise before, for example Hubrich (2005). However, we contribute to the literature by using more recent data which includes both the financial crisis as well as the European sovereign debt crisis, and by finer disaggregation of Euro Area HICP into 13 sub-components for all 17 Euro Area countries.

This paper is organized as follows: Section 2 goes through the theory underlying inflation dynamics through an historic perspective on the Phillips curve. Section 3 covers previous research regarding forecasting Euro Area HICP using sub-components. Section 4 goes through some standard estimation procedures in

time series econometrics. Section 5 describes the modeling strategy and covers the different models. Section 6 describes the data set. Section 7 demonstrates the use and the different issues that arise when using univariate and Phillips curve based models when forecasting inflation. In section 8, we run the exercise of seeing if forecasting the sub-components of HICP outperforms the direct forecast. Section 9 concludes.

2 The Phillips Curve

In this section, we will briefly go through the theory underpinning inflation dynamics. The main framework to understand inflation dynamics today is the Phillips curve. The Phillips curve has gone through several iterations and we find it instructive to understand inflation dynamics by briefly going through the history of the Phillips curve. With the resurgence of Keynesian thought, you have a case in point that you shall not underestimate old knowledge and the light it might bring on newer ideas and the understanding of how those newer ideas where formed. For a more comprehensive history of the development of the Phillips curve, we refer to Gordon (2011) and Fuhrer et al. (2009).

It started with Phillips (1958) noticing that a higher (lower) unemployment rate was related to a lower (higher) rate of change in nominal wages in United Kingdom during 1861-1957. Phillips made several scatter diagrams to show this relationship, see figure 2.1 for an example. Phillips got the solid line in the figure by fitting the equation:

$$\triangle w_t + a = bu_t^c \Leftrightarrow \log_{10}(\triangle w_t + a) = \log_{10}(b) + c\log_{10}(u_t)$$

where $\triangle$ is the difference operator (i.e.$\triangle x_t = x_t - x_{t-1}$), w_t is money wage rates at time t, u_t is the unemployment rate, a,b and c are constants.

Phillips estimated this equation in the following way:

First he calculated averages of $\triangle w_t$ where u_t was in the interval 0-2, 2-3, 3-4, 4-5, 5-7 and 7-11, this is represented by the crosses in figure 2.1.

Then he estimated b and c by the least squares method using the values of the first four crosses and decided a by trial and error to make the curve fit as good as possible to the remaining two crosses.

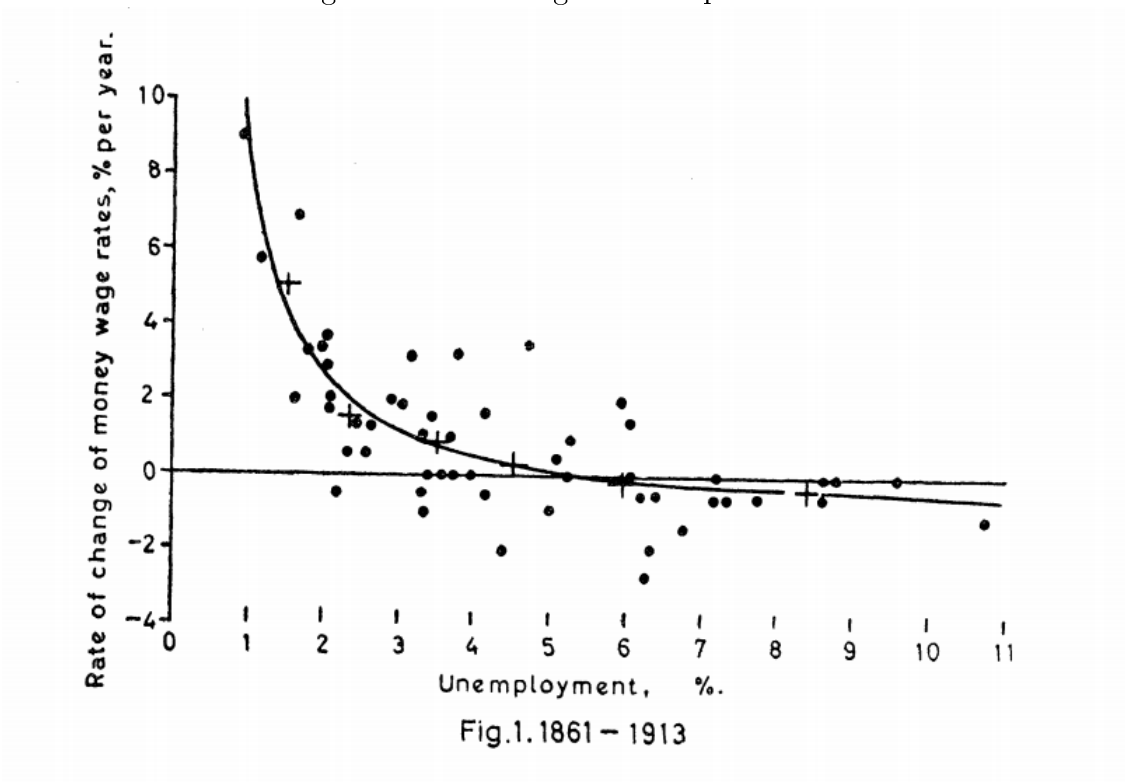
He arrived at the following values:

$$\triangle w_t + 0.9 = 9.638u_t^{-1.394} \Leftrightarrow \log_{10}(\triangle w_t + 0.9) = 0.984 - 1.394\log_{10}(u_t)$$

The interpretation of this model is as follows: If the unemployment rate was approximately 5.479 we would have zero wage growth. A higher unemployment

rate would lead to decreasing wages and lower unemployment rate would lead to increasing wages.

Figure 2.1: The original Phillips curve



Source: Phillips (1958)

What is less known is that Phillips also discussed a number of wage determinants, such as increased import prices which would have an effect on the cost of living and be a factor in wage negotiations, which later have received a lot of attention in the literature of wage and price determination. Furthermore of interest is also that Phillips suggested the possibility of a "speed limit" effect, that not only the level but also the rate of change has important consequences for the change in nominal wages. Phillips noted also the reluctance of workers to accept nominal wage cuts which would suggest that the Phillips curve could be non-linear.

Samuelson and Solow (1960) popularized the name Phillips curve and explored its policy implications. They note that in the first Phillips curve there is a trade-off between inflation and unemployment so theoretically policy makers could choose a pair which they found most optimal. However they argue that if any such trade-

off existed it must only be in the short-run and could shift the Phillips curve. Due to their failure to discuss the long-run in more detail than they did they consequentially became criticized for posing a long-run inflation-unemployment tradeoff available for exploitation by policymakers. Personally from reading their article it seems a bit harsh, rather their short-run was medium-run and presented a trade-off only for a few years.

Friedman (1968) and Phelps (1967, 1968) are both credited to discover the natural rate hypothesis. Friedman argued that monetary policy could only lower the unemployment rate temporarily, the mechanics being that a lower interest rate stimulates spending, raise prices, raise the marginal products of labor, and increases employment and output. Friedman believed prices would rise before wages, lowering the real wage received; thereby prompting increased nominal wages demand by labor and ultimately wage increases would match accumulated price increases and bring unemployment back to its natural rate. Phelps argued that the Phillips curve shifts uniformly upward one point for every one point increase in inflation expectations. Using an adaptive expectations framework, workers expected inflation to be the same as it has been in the past and Phelps developed the accelerationist Phillips curve:

$$\pi_t = \pi_t^e - \lambda u_t = \pi_{t-1} - \lambda u_t \Longleftrightarrow \triangle \pi_t = -\lambda u_t,$$

where π_t is the inflation rate at time t, π_t^e is the expected inflation rate at time t, the coefficient$-\lambda$ measures the slope of the Phillips curve. With a small adjustment to the accelerationist Phillips curve we can get the Non-Accelerating Inflation Rate of Unemployment (NAIRU):

$$\triangle \pi_t = -\lambda(u_t - u^N),$$

where u^N is the natural rate of unemployment.

We now have arrived at the textbook NAIRU model which says that when the unemployment rate is below the natural rate the economy experiences inflationary pressures and when the unemployment rate is above the natural rate the economy experiences deflationary pressures and we reach a stable inflation rate at the natural unemployment rate. The NAIRU offers two important insights, first that

there is no long-term a tradeoff between inflation and unemployment. The second is the role expectations have in the price-setting process, which became a huge component in further developing inflation models and is still today.

Muth (1961), the father of rational expectations theory, noted that economists used ad-hoc exogenous equations for describing the mechanics of expectations, Muth wanted more consistency as in that the expectations should be formed from the prediction of the economic theory, i.e. to make expectations endogenous within the model. Lucas (1972, 1973) and Sargent and Wallace (1975) developed models building upon rational expectations. What they found in their models were that the price level based on rational expectations was extremely flexible, that the only effect monetary policy could have was when it shifted the money supply unanticipated, making monetary policy practically inefficient and business cycles obsolete. So it's not surprising their models fail the empirical tests, however their work is important and lays an foundation for further work by Fischer (1977), Gray (1977), Taylor (1980), Calvo (1983) and others that emphasized staggered nominal wage and price setting by forward looking individuals and firms. Wage and price rigidities made monetary policy valid again inside the rational expectations framework. This work leads to the New Keynesian Phillips Curve (NKPC) which is basically a forward looking Phillips curve:

$$\pi_t = E_t(\pi_{t+1}) - \lambda(u_t - u^N),$$

where $E_t()$ is the conditional expectation given data up to time t.

An interesting feature with the NKPC is that as it's only forward looking, it would be possible to achieve low inflation immediately without the increase in unemployment by simply changing expectations. This is very interesting as it relates to the value of monetary policy credibility, anchoring of inflation expectations and forward guidance which are quite hot topics today. Gordon (2011) argues that the NKPC has its application to economies with unstable macroeconomic environment, like the four hyperinflations Sargent (1982) studied using the NKPC. The problem with the NKPC despite it having some nice theoretical underpinnings to it is that it fails the empirical test. Data shows that inflation is very persistent (see Fuhrer and Moore (1995)) and the NKPC has troubles generating that degree of persistence.

This article is about forecasting and the Phillips curve we consider is not the NKPC, the main workhorse model for forecasting inflation, particular at central banks is Gordon's (1977) "triangle model" which is also the model that forms our basis. The story behind the "triangle model" is that it tried to explain the 1970s stagflation, which saw that a sharp increase in oil-prices which lead to both inflation and higher unemployment. So basically you just introduce a "supply-shock" term into the NAIRU and you get the "triangle model":

$$\pi_t = \pi_{t-1} - \lambda(u_t - u^N) + z_t,$$

where z_t is the supply shock. It's called the "triangle model" as it has three drivers, built-in inflation (from inflation expectations and the fact inflation is persistent), demand-pull inflation (the output gap) and cost-push inflation (the supply shocks).

By going through some of the history of the Phillips curve, we see that inflation has many dynamic determinants such as inflation expectations through past inflation experience, wage negotiation power through unemployment level, import prices through exchange rates and etc. These relationships do not necessarily have to remain stable as they are affected by individuals' behavior, institutions such as labor unions, monetary policy, fiscal policy, and more which in turn could react to changes in outcome, case in point example being Deutsche Bundesbank strong inheritance from the hyperinflation period after the Second World War.

With this we wish to warn that one cannot conclude that the "triangle model" is more "true" than the NKPC just because the "triangle model" explains the historical data better. It depends on how you wish to use the model, for example is it for forecasting or policy evaluation. The "triangle model" has outdone the NKPC in forecasting so far in the literature, while the NKPC can explain why "forward guidance" and "credibility" has been such important issues for central banks lately. Lucas summarizes this nicely in his famous Lucas Critique (1976): "... features which lead to success in short-term forecasting are unrelated to quantitative policy evaluation, that the major econometric models are (well) designed to perform the former task only, and that simulations using these models can, in principle, provide *no* useful information as to the actual consequences of alternative economic policies. These contentions will be based not on deviations between

estimated and "true" structure prior to policy change but on the deviations between the prior "true" structure and the "true" structure prevailing afterwards". It makes you think how one in today's environment (of high unemployment rate and super active central banks in the advanced economics) best would forecast inflation in the medium term (perhaps even in the short term). Perhaps it's not then surprising that there are some divergence, those who forecast deflation and those who forecast hyperinflation, depending on the models they choose to emphasize.

3 Previous Research

In this section, we will go through some of the papers that focus on forecasting Euro Area inflation with the use of sub-components. If you do not have much experience with econometrics we would recommend that you read section 5 before this section since it cannot really be helped that a lot of terminology and concepts will be used in this section without going into the details.

There are several papers which try to improve on inflation forecasts by using sub-components. See for example Aron and Muellbauer (2008) for the USA case and Demers and De Champlain (2005) for the Canadian case. However, as we focus on Euro Area inflation we will mainly discuss the papers that also have the Euro Area as there focus and especially those papers with comparable methodology which we can compare our results to. Unfortunately, we have not been able to find any study which gives an overview nor any cross-country study (if you don't count the Euro Area) that tries to test if the results from using sub-components in inflation forecasting can be generalized to most countries.

The results from the literature is mixed, some find that using sub-components information can help forecast performance while some find it doesn't except for the very short time-horizon, 1 to 3 month ahead. It seems to matter if you modeled inflation using M-o-M or Y-o-Y and also if you used quarterly or monthly series. Different papers have different time period for the data which makes comparison harder.

Marcellino, Stock, and Watson (2003) study the forecasting performance for several variables including inflation rate using country-specific data for the Euro Area. They use several models: univariate autoregresssions (AR), vector autoregressions (VAR), single equation models as well as a dynamic factor model (DFM) which work by measuring the co-movements of multiple series and taking them out as factors which can be regressed on the variable of interest. Their data is monthly and covers 1982 to 1997. They find that no multivariate models beats their pooled univariate autoregressions, also their DFM outperformed the VAR's. So they saw gains by forecasting at the country-level and then aggregating than to directly forecast at the Euro Area level, their relative RMSE compared to the direct forecast is 0.90.

Hubrich (2005) finds that forecasting sub-components which she disaggregates to five sub-components, services, goods, processed food, unprocessed food and energy, doesn't help in forecasting the HICP but may slightly help in forecasting HICPX. Hubrich tries several forecast models (random-walk, AR, VAR), different model selection procedures, to forecast both HICP and HICPX (core inflation) using the sample 1992m1-2001m1. She finds that aggregating helps for the one-month ahead forecast of the Y-o-Y HICP but performs worse 6-months and 12-months ahead, for HICPX aggregating helps if only slightly for all three horizons.

Benalal et al. (2004) basically arrive at the same results as Hubrich (2005) using pretty much the same models and data (1990m1-2002m6). However, Den Reijer and Vlaar (2003) and Espasa and Albacete (2004) get opposing results. Both find that forecasting the disaggregates and combining them outperform the direct forecast on the aggregate on all time-horizons from 1-18 months forward for the Y-o-Y inflation rate. However, one reason why their results may differ is that both make use of a vector error correction model (VECM) which Hubrich and the others didn't.

The theoretical motivation for working with sub-aggregates is discussed in Hendry and Hubrich (2010). They study whether it's better to combine disaggregates forecasts or to include disaggregate information to forecast an aggregate or just simply use the aggregate only. They derive analytical results for the case when the data-generation process (DGP) is affected by a changing coefficient, miss-specification, estimation uncertainty and miss-measurement error. A structural break at the forecast origin affect absolute, but not relative, forecast accuracies; miss-specification and estimation uncertainty induce forecast-error differences, which variable-selection procedures or dimension reductions can mitigate. They also perform Monte Carlo simulations to test their analytic results for changing coefficient, miss-specification and miss-measurement error, in which they conclude that adding disaggregate information when forecasting the aggregate is the best approach, i.e. there exists valuable information in the disaggregates but it's better to incorporate in a model with the aggregate than to forecast the sub-aggregates and combine them to the aggregate. They also did an empirical study, but for US inflation. However, none of their models could outperform the direct AR forecast. Interestingly they found that modeling inflation in M-o-M changes

and then evaluated at Y-o-Y gives more accurate forecasts than working with the Y-o-Y series directly as we do in this paper as does most other papers as well. Also, no qualitative difference was observed between working with the level of inflation or the change in inflation.

4 Standard Estimation Procedures

In this paper we use two models, the autoregressive moving-average (ARMA) model and the Autoregressive Distributed Lag (ADL) model.

ARMA: The general ARMA model can be defined as follows:

$$\begin{cases} y_t = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + x_t \\ x_t = \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i} + \varepsilon_t \end{cases} \qquad (ARMA(p,q)),$$

where α is a constant (the intercept), $\{y_t\}$ is the ARMA-process, $\{x_t\}$ is the MA part, p is the order of lags for the AR-part, q is the order of lag for the MA part and ε_t is white noise, also called residuals in an fitted model.

The sequence $\{\varepsilon_t\}$ is a white noise process if for each time t satisfies the three following properties:

1. Mean of zero, $E(\varepsilon_t) = E(\varepsilon_{t-1}) = \cdots = 0$
2. Constant variance, $E(\varepsilon_t^2) = E(\varepsilon_{t-1}^2) = \cdots = \sigma^2$
3. Serially uncorrelated, $E(\varepsilon_t \varepsilon_{t-s}) = E(\varepsilon_{t-j}\varepsilon_{t-j-s}) = 0, \ for\ all\ s$

$\{x_t\}$ will be stationary if $\sum_{i=1}^{q} \gamma_i^2$ and $(\gamma_s + \gamma_1\gamma_{s+1} + \gamma_2\gamma_{s+2} + \ldots)$ both are finite. $\{y_t\}$ will be stationary as long as the roots of $1 - \sum_{i=1}^{p} \beta_i z^p = 0$ lies outside the unit circle.

For normal statistical inference it's enough for $\{x_t\}$ and $\{y_t\}$ to be weakly stationary, or also called covariance-stationary. A stochastic process, $\{y_t\}$ having a finite mean and variance is covariance-stationary if for all t and $t-s$, we have:

1. Constant mean, $E(y_t) = E(y_{t-s}) = \mu$
2. Constant variance, $E((y_t - \mu)^2) = E((y_{t-s} - \mu)^2) = \sigma_y^2$
3. Constant covariance, $E((y_t - \mu)(y_{t-s} - \mu)) = E((y_{t-j} - \mu)(y_{t-j-s} - \mu)) = \rho_s$

For estimation, if we have no MA-terms, ordinary least squares (OLS) can be used. In OLS estimation we simply choose the parameters which minimize the sum of squared residuals. So in the AR(1) case:

$$y_t = \alpha + \beta_1 y_{t-1} + \varepsilon_t,$$

where we have T observations of y_t, so we wish to minimize $\sum_{t=1}^{T} \varepsilon_t^2 = \sum_{t=0}^{T} (y_t - \alpha - \beta_1 y_{t-1})^2 = f(\alpha, \beta_1)$ with respect to both α and β_1. Taking first derivatives and setting to zero, we get:

$$\begin{cases} f'_\alpha(\alpha, \beta_1) = -2\sum_{t=1}^{T} (y_t - \alpha - \beta_1 y_{t-1}) = 0 & (1) \\ f'_{\beta_1}(\alpha, \beta_1) = -2\sum_{t=1}^{T} y_{t-1} (y_t - \alpha - \beta_1 y_{t-1}) = 0 & (2) \end{cases}$$

Now all we need to do is solve (1) and (2). Let us start with (1), after dividing by -2 and using $\sum_{t=1}^{T} y_t = T\bar{y}_t$, where $\bar{y}_t$ is the average of y_t, we get that $\sum_{t=1}^{T} (y_t - \alpha - \beta_1 y_{t-1}) = T\bar{y}_t - T\alpha - T\beta_1 \bar{y}_{t-1} = 0$, and solving for α :

$$\alpha = \bar{y}_t - \beta_1 \bar{y}_{t-1} \qquad (3)$$

We now substitute (3) into (2) and again divide by -2 and expand the sum to get:
$\sum_{t=1}^{T} y_{t-1} y_t - \sum_{t=1}^{T} y_{t-1} \bar{y}_t - \sum_{t=1}^{T} y_{t-1} \beta_1 \bar{y}_{t-1} - \sum_{t=1}^{T} y_{t-1} \beta_1 y_{t-1} = 0$ and solving for β_1 we get:

$$\beta_1 = \frac{\sum_{t=1}^{T} y_{t-1} y_t - T\bar{y}_{t-1}\bar{y}_t}{\sum_{t=1}^{T} y_{t-1}^2 - T\bar{y}_{t-1}^2} =$$

$$= \frac{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1})(y_t - \bar{y}_t)}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1})^2} \qquad (4)$$

One can easily verify that the right-hand side of (4) equals the left-hand side by expanding the parenthesis:

$$\frac{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1})(y_t - \bar{y}_t)}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{t-1})^2} =$$

$$= \frac{\sum_{t=1}^{T} y_{t-1}y_t - \sum_{t=1}^{T} y_{t-1}\bar{y}_t - \sum_{t=1}^{T} \bar{y}_{t-1}y_t + \sum_{t=1}^{T} \bar{y}_{t-1}\bar{y}_t}{\sum_{t=1}^{T} y_{t-1}^2 - 2\sum_{t=1}^{T} y_{t-1}\bar{y}_{t-1} + \sum_{t=1}^{T} \bar{y}_{t-1}^2} =$$

$$= \frac{\sum_{t=1}^{T} y_{t-1}y_t - T\bar{y}_{t-1}\bar{y}_t - T\bar{y}_{t-1}\bar{y}_t + T\bar{y}_{t-1}\bar{y}_t}{\sum_{t=1}^{T} y_{t-1}^2 - 2T\bar{y}_{t-1}^2 + T\bar{y}_{t-1}^2}.$$

When MA-terms are included, we can no longer use OLS as we don't directly observe $\{\varepsilon_t\}$ sequence. A common way to estimate the coefficients is by using maximum likelihood estimation (MLE).

Assuming that $\{\varepsilon_t\}$ are drawn from a normal distribution having a mean of zero and a constant variance of σ^2, we get from standard distribution theory that the likelihood of any realization of ε_t is:

$$L_t = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(\frac{-\varepsilon_t^2}{2\sigma^2}\right),$$

where L_t is the likelihood of ε_t.
Since the realizations of $\{\varepsilon_t\}$ are independent, the likelihood of the joint realizations is the product of the individual likelihoods. Hence, we get:

$$L = \prod_{i=1}^{T} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) exp\left(\frac{-\varepsilon_t^2}{2\sigma^2}\right)$$

And taking the natural logarithm on both sides we get the log likelihood:

$$\ln(L) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{T} \varepsilon_t^2.$$

The procedure used in maximum-likelihood estimation is to select the distributional parameters so as to maximize the likelihood of drawing the observed sample. So if we want to estimate the MA(1) model:

$$x_t = \gamma_1\varepsilon_{t-1} + \varepsilon_t,$$

which we can also write as

$$\varepsilon_t = x_t - \gamma_1\varepsilon_{t-1} = x_t - \gamma_1 L\varepsilon_t,$$

where $L\varepsilon_t$ is the lag operator ($L\varepsilon_t = \varepsilon_{t-1}$). So we can solve for ε_t:

$$\varepsilon_t = \frac{x_t}{1+\gamma_1 L} = \sum_{i=0}^{t-1}(-\gamma_1)^i x_{t-i},$$

which will be a convergent process as long as $\|\gamma_1\| < 1$. We can then substitute this into our log likelihood and arrive at:

$$\ln(L) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(\sum_{i=0}^{t-1}(-\gamma_1)^i x_{t-i}\right)$$

which we maximize by adjusting σ^2 and γ_1. However, note that we won't get as simple first order conditions like we did with the AR(1) model. Numerical optimization routines are used to find the values of σ^2 and γ_1.

ADL: The general ADL model can be defined as the following:

$$y_t = \alpha + \sum_{i=1}^{p}\beta_i y_{t-i} + \sum_{i=0}^{q_1}\gamma_{1,i}x_{1,t-i} + \sum_{i=0}^{q_2}\gamma_{2,i}x_{2,t-i} + \cdots + \sum_{i=0}^{q_k}\gamma_{k,i}x_{k,t-i} + \varepsilon_t$$

where $\{y_t\}$ and $\{x_t\}$ are stationary processes, and ε_t is white noise. As long as $\{\varepsilon_t\}$ is a white noise process the ADL model can be estimated using OLS.

Estimation: We used the statistical software package EViews[1] to aid us in estimating the coefficients to our models and all other computations. However, EViews does not solve the models analytically, instead EViews uses nonlinear regressions techniques when estimating both ARMA and ADL models, the reason why they use nonlinear techniques even though the parameters are linear is that it has the advantage of being easy to understand, generally applicable, and easily extended to nonlinear specifications and models that contain endogenous right-hand side variables. I.e. EViews generalizes the models and then solve them numerical.

[1]See http://www.eviews.com for further information about the statistical software package.

5 Forecasting Methodologies

This section has three subsections. In subsection 5.1 we go through some forecasting terminology and typical data transformation used in the literature. In subsection 5.2 we will go through the forecasting models used in this paper. In subsection 5.3 we will describe the forecast methods and model selection procedure of the different models.

5.1 Terminology

h-period inflation: We denote h-period inflation by $\pi_t^h = h^{-1}\sum_{i=0}^{h-1}\pi_{t-i}$, where π_t is the monthly rate of inflation at an annual rate, i.e. $\pi_t = 1200\ln(P_t/P_{t-1})$, where P_t is the price index at time t and ln stands for the natural logarithm. The log transformation is simply used because it allows us to arithmetically add instead of multiplying the inflation rates, so the one-year ahead year-on-year (Y-o-Y) inflation rate is given by $\pi_{t+12}^{12} = 12^{-1}\sum_{i=0}^{12-1}\pi_{t+12-i} = 100\ln(P_{t+12}/P_t)$.

Direct and iterated forecast: There are two ways to make an h-period ahead forecast model. First the direct way is to regress π_t^h on t-dated variables (variables observed at time t). The second way is the iterated forecast which builds on the one-step ahead model. For example π_{t+1} is simply regressed on just π_t, which is then iterated forward to compute future conditional means, i.e. if we assume our model is given by $\pi_t = \beta\pi_{t-1} + \varepsilon_t$ then the two-step ahead forecast will be $E_t(\pi_{t+2}) = \beta E_t(\pi_{t+1}) = \beta^2\pi_t$, where $E_t()$ is the conditional expectation given data up to time t and $E_t(\varepsilon_s) = 0$ for all $s \geq t$. If predictors other than past π_t are used, then this requires subsidiary models for the predictor, or alternatively, modeling π_t and the predictor jointly, for example as a vector autoregression (VAR) and iterating the joint model forward.

Pseudo out-of-sample forecasts - rolling and recursive estimation: Pseudo out-of-sample forecasting simulates the experience of a real-time forecaster by performing all model specification and estimation using data through date t, making a h-step ahead forecast for date $t+h$, then moving forward to date $t+1$ and

repeating this through the sample.[2] Pseudo out-of-sample forecast evaluation captures model specification uncertainty, model instability, and estimation uncertainty, in addition to the usual uncertainty of future events. Model estimation can either be rolling (using a moving window of fixed size) or recursive (using an increasing window, always starting with the same observation). Rolling estimation is preferred if one believes that the data-generating process (DGP) has changed over time, i.e. the data exhibits structural change. This is because using early estimates from an earlier DGP would bias the parameter estimates of the current DGP and lead to a biased forecast. However, there is a trade-off, by reducing the sample one increases the variance in the parameter estimates and therefore also the forecast errors. It is worth noting that Stock and Watson (1996) shows that most macroeconomic series does exhibit structural change, so if one models a longer time period it is important to either directly model the structural change or allow the parameters to change over time which can be done by rolling estimation. We will be using both rolling and recursive estimation as we won't directly model structural change and have limited sample size to both estimate a lot of parameters and produce long enough out-of-sample forecasts.

Dependent vs. independent variables: The "dependent variable" represents the output or effect, or is tested to see if it is the effect. In our case, the dependent variable is the Y-o-Y inflation rate. The "independent variables" represent the inputs or causes, or are tested to see if they are the cause. We will be using several variables as "independent variables", for example the unemployment rate. Note also that common synonyms for "independent variables" are "regressor", "controlled variable" and "explanatory variable".

Statistically significance, t-test, test statistic, null-hypothesis and p-value: In econometrics, you often want to test if an independent variable is statistically significant, i.e. if the variable show some kind of pattern with the dependent variable which is not just by chance. This is done by a t-test which we now will go through. We start with the t-statistic, which follows the Student's t

[2] A strict interpretation of pseudo out-of-sample forecasting would entail the use of real-time data (data of different vintages), but we interpret the term more generously to include the use of final data.

distribution if the null-hypothesis is supported, the t-statistic is defined as:

$$t_{statistic} = \frac{\beta - \beta_{null}}{\sigma_\beta},$$

where β is the estimated coefficient to the independent variable, β_{null} is the value of β under the null-hypothesis and σ_β is the sample standard deviation of β.

As we want to test if the independent variable is statistically significant we have that $H_0 : \beta_{null} = 0$, $H_1 : \beta_{null} \neq 0$, where H_0 and H_a stands for the null-hypothesis and the alternative-hypothesis respectively. We now can use the p-value which is the probability of obtaining a test statistic at least as extreme as the one that was actually observed. To demonstrate, say that we estimated $\beta = 1$ and the standard error of β to be $\sigma_\beta = 0.5$, under the null-hypothesis $H_0 : \beta_{null} = 0$ we get the t-statistic, $t_{statistic} = 2$. If we have a large sample, the Student's t distribution is well approximated by the normal distribution so we can calculate the p-value as follows:

$$p - value = 2\Phi\left(-\left\|t_{statistic}\right\|\right),$$

where Φ is the standard normal probability density function.

In our case the we have the $p - value = 0.046$, which means the probability of obtaining $\beta = 1$ if the true value is $\beta = 0$ is only 4.6% so we can reject the null-hypothesis, $H_0 : \beta_{null} = 0$, i.e. that the independent variable is not statistically significant with significance level of 4.6%. Common significance level used are 10%, 5% and 1%, so if your test statistic is so extreme that the probability of obtaining it is less than 10%, 5% or 1% depending on the significance level you can reject the null-hypothesis.

***F*-test:** In econometrics, the F-test is used for testing joint null hypothesis, for example that all unemployment lags are statistically insignificance or that all variables are insignificant. It can also be used as a test for additional information, say you have a set of explaining variables and you want to test if adding another significantly improves the fit to the data, i.e. that it contains additional information which wasn't in the previous set of explaining variables.

As an example, we have the model:

$$\pi_t = \beta_0 + \beta_1 \pi_{t-1} + \beta_2 u_{t-1} + \varepsilon_t,$$

and we want to test if there is any informational content in past inflation and unemployment rate for explaining the current inflation rate. I.e. our null-hypothesis and alternative hypothesis is $H_0 : \beta_1 = \beta_2 = 0$, $H_1 : \beta_1 \neq 0 \, and/or \, \beta_2 \neq 0$ respectively. You can view this as we have two models, the "unrestricted" one above, and the "restricted" one with only the constant, β_0, as explaining variable. The F-statistic is given by:

$$F_{stastic} = \frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{T - p_2}},$$

where T is the number of observations, RSS_i is the residual sum of squares of model i, i.e. $RSS = \sum_{t=1}^{T} \varepsilon_t^2$ and p_i is the number of parameters in model i, i.e. in our example: $p_1 = 3$, $p_2 = 1$.

The F-statistic follows an F distribution with $(p_2 - p_1, T - p_2)$ degrees of freedom, and we can again calculate a p-value to see if we can reject the null-hypothesis or not.

Akaike Information Criterion (AIC): Is a common model selection criterion which is defined as:

$$AIC = T \ln \left(\sum_{t=1}^{T} \varepsilon_t^2 \right) + 2n,$$

where n = number of parameters estimated (p + q + possible constant term if an ARMA(p,q) model), T = number of usable observations. The AIC rewards models which have better fit (lower sum of squared residuals) and punishes models with requires many parameters to be estimated so we end up with an parsimonious model.

Ljung-Box test: Is used to check if a process behaves like a white noise process (in which all autocorrelations should be zero). The Ljung-Box calculates the Q-statistic as:

$$Q = T(T+2)\sum_{k=1}^{s}\frac{r_k}{T-k},$$

where T is the sample size, r_k is the sample autocorrelation at lag k and s is the number of lags being tested. If the sample value of Q exceeds the critical value of a chi-squared distribution with s degrees of freedom, then at least one value of r_k is statistically different from zero at the specified significance level.

The sample autocorrelation, r_k is calculated as:

$$r_k = \frac{\sum_{t=s+1}^{T}(y_t-\bar{y})(y_{t-s}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2},$$

where $\bar{y} = \sum_{t=1}^{T} y_t$, i.e. the sample average.

General-to-Specific modeling: Is when you usually start with an over-parameterized model and then follow a parameter reduction strategy. This can be done manually by for example examining correlograms (also known as an autocorrelation plot), estimate different models and test the coefficients with t-test and F-tests. There also exist several algorithms some very advanced which automate the procedure. These algorithms usually follow four steps: First check that the model is well-behaving. Second, remove a variable or variables that satisfy the selection criteria. Third, check if the model is still well-behaving. Fourth, continue doing the second and third step until no further variables can be removed by the selection criteria. For a more comprehensive overview on the general-to-specific modeling approach, we refer to Campos, Ericsson and Hendry (2005). In this paper we will develop our own algorithms which are very simple and mostly only relies on t-test and F-test as well as the Akaike information criterion (AIC). We will describe these algorithms in more detail in subsection 5.3.

Root mean squared error (RMSE) and rolling RMSE: RMSE is a measure of the forecast performance, the RMSE of the h-period ahead forecasts made over the period t_1 to t_2 is

$$RMSE_{t_1,t_2} = \sqrt{\frac{1}{t_1 - t_2 + 1} \sum_{t=t_1}^{t_2} (\pi_{t+h}^h - E_t(\pi_{t+h}^h))^2},$$

where $E_t(\pi_{t+h}^h)$ is the pseudo out-of-sample forecast of π_{t+h}^h made using data through date t.

In this paper, we often use the relative RMSE, which is the model's RMSE divided by the RMSE from the naive forecast. We also use a rolling RMSE, which is computed using a weighted centered 25-month window:

$$rollingRMSE(t) = \sqrt{\sum_{s=t-12}^{t+12} K(|s-t|/13)(\pi_{s+h}^h - E_s(\pi_{t+h}^h))^2 / \sum_{s=t-12}^{t+12} K(|s-t|/13)},$$

where K is the bi-weight kernel, $K(x) = (15/16)(1-x^2)^2 \mathbf{1}(|x| \leq 1)$, see Stock and Watson (2008). Moreover, this kernel puts more weight for the center, so the rolling RMSE is basically a central moving average with less weight on the tails, which produces smooth graphs which makes comparing two different models for different time periods a lot more easier.

5.2 Forecasting Models

Autoregressive–moving-average (ARMA): We will only use one specification for the ARMA where we use the Y-o-Y inflation rate:

$$\pi_t^{12} = \alpha + \sum_{i=1}^{p} \beta_i \pi_{t-i}^{12} + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i} + \varepsilon_t \qquad (ARMA(p,q))$$

Phillips curve based models: Are generally models that include the unemployment rate (or another explanatory variable to proxy economic activity) as well as past inflation rate as explanatory variables for future inflation rate.

We will consider eight general model specifications. The reason why so many is that we want to be able to answer three methodological questions, given the application on using Phillips curve based models to forecast the one-year ahead Y-o-Y Euro Area HICP:

1. Is it better to use a direct forecast or iterated forecast?
2. Should one use past year-over-year (Y-o-Y) or month-over-month (M-o-M) inflation rate as explanatory variables?
3. Should one model the Y-o-Y inflation rate as a unit-process or not?

$$\pi_{t+12}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF1)$$

$$\pi_{t+12}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF2)$$

$$\pi_{t+12}^{12} - \pi_t = \alpha + \sum_{i=0}^{L_1} \beta_i \triangle \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF3)$$

$$\pi_{t+12}^{12} - \pi_t = \alpha + \sum_{i=0}^{L_1} \beta_i \triangle \pi_{t-i} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF4)$$

$$\pi_{t+1}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF1)$$

$$\pi_{t+1}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF2)$$

$$\pi_{t+1}^{12} - \pi_t = \alpha + \sum_{i=0}^{L_1} \beta_i \triangle \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF3)$$

$$\pi_{t+1}^{12} - \pi_t = \alpha + \sum_{i=0}^{L_1} \beta_i \triangle \pi_{t-i} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF4)$$

Where DF stands for direct forecast and IF stands for iterated forecast, π_{t+12}^{12} is the Y-o-Y inflation rate one-year ahead, π_t^{12} is the Y-o-Y inflation rate at time t, u_t is the unemployment rate at time t, z_i's are the "supply shock" variables in Gordon's "triangle model", however we define z more generally as a variable which may help predict the inflation rate. D_i are dummy variables for each month to account for seasonality. $L_1$is the number of lags for past inflation rates, L_2 is the number of lags in unemployment rates and $L_{3,i}$ is the number of lags for the i'th "supply shock" variable.

Hence we have four direct forecasts and four iterated forecasts. For each type, we also have two that doesn't model the Y-o-Y inflation rate as a unit-process and two that does and the difference between these two is that one uses Y-o-Y and the other uses M-o-M inflation rates as past inflation. By comparing the performance of these model specifications our goal is to be able to answer the methodological questions we stated earlier. Variants of these models can be found in the literature, all models used by Marcellino, Stock, and Watson (2003) are in the form of DF1 without the seasonal dummies. Stock and Watson (1999) uses IF4 but without the seasonal dummies. Benalal et al. (2004) also uses seasonal dummies in their specifications, but they use M-o-M inflation rates instead of Y-o-Y. The DF models are empirical versions of Gordon's "triangle model" which we covered in Section 2, the difference being that we allow for

several lags, included seasonal dummies and some versions which don't require the inflation rate to be a unit-root process and allow for more freedom in what counts as "supply variables". The IF models are more similar to the NAIRU as they do not have any "supply shock" variables and with the same generalizations to become empirical models. The reason why none of the IF models include any "supply shock" variables is because then one would need to forecast those which often one can't do very satisfying, and Stock and Watson (1999) notes that "supply shock" variables are statistically significant in full-sample specifications but produce worse out-of-sample forecasts when included. In the DF's we are satisfied with using u_t, instead of the contemporaneous unemployment rate, u_{t+1}. However, for the IF's we use u_{t+1} as we anyways need a subsidiary forecast of u_t.

An interesting implication from our specifications which models Y-o-Y inflation rate as a unit root, i.e. DF3, DF4, IF3 and IF4 is that they imply a constant NAIRU, you can see this from that the NAIRU is $\triangle\pi_t = -\lambda(u_t - u^N)$, so $\alpha + \gamma_0 u_t = -\lambda(u_t - u^N) \implies \alpha = -\gamma_0 u^N \implies u^N = -\alpha/\gamma_0$.

It's important to note that we have not explicitly modeled any structural breaks, or for example allowed for time-varying parameters (see Staiger, Stock and Watson (1997) for the first time-varying NAIRU model). However, as our approach is to always re-estimate the parameters for each window and judge the performance by pseudo out-of-sample forecasts we believe it's not necessarily to explicitly model structural breaks.

5.3 Forecast Procedures

Our main goal is to find the model which performs best in forecasting, which we measure by the RMSE for the out-of-sample forecast. So, we wish to simulate the real-time forecaster experience as close as possible. The general way we do this is to start with a first estimation window of 5 years, a shorter window gave too few observations for an accurate first estimation and using a longer window reduces our out-of-sample period which we use to judge our model with. We then try several models from a family of models and use the best in-sample fit for the current window and forecast ahead one-year. We then recursively grow the window with one period and then again select a model from the family of models with the best in-sample fit, and forecast ahead one-year. We then finally calculate the RMSE from the produced forecast and actual values, and this is the sole criteria we evaluate our forecasts on. Below we will go through a little bit more specific of the choices we make in the forecast procedures for both the univariate and Phillips curve based models.

ARMA: The common method of working with an ARMA model is to use the Box-Jenkins methodology, see Box and Jenkins (1976).

The Box and Jenkins methodology consist of three stages:

1. **Identification stage:** Visualize the data to check for outliers, missing values, structural breaks, seasonality, non-stationarity and etc. Also graph the correlograms to give an idea which AR or MA terms should be included. We use an ADF-test to check for non-stationarity.

2. **Estimation stage:** Estimate and examine the model coefficients, the goal is to select a stationary and parsimonious model.

3. **Diagnostic check:** Check that the residuals behave as a white-noise process. We employ the Ljung-Box test and throw away those models that fail the test.

However, since we wish to automate the model selection as we will be running multiple forecasts we develop two algorithms which are based on the principles from Box-Jenkins and hopefully select a similar model as if we used Box-Jenkins.

In the first algorithm we start with the 5 year window and estimate all ARMA models and remove those that fail the diagnostic test and select the model that gives the lowest AIC. We then expand the window one step and re-estimate all ARMA models, we call this AIC. For the second algorithm we use the same ARMA model for each step, we do this for all ARMA models and choose the ARMA model which gives the lowest RMSE, we call this ARMA. In the second algorithm we use the full sample for model selection and therefore cheat a little on the true real-time forecaster experience, however if a parsimonious model is used and as the parameter estimation uses only the data up to t, we believe to still concur with the real-time forecaster experience.

Phillips curve based models: The procedure is similar for the Phillips curve based models, the same growing recursive window estimation and forecasting. What differs is that we do not do any diagnostic check on the residuals; one reason for this is that in the direct forecast you cannot expect that the residuals behave as white-noise as the Y-o-Y inflation rate is highly persistent and so will the forecast error also be. What also differs is the model selection procedure, here we don't only rely on AIC; we now reduce the model with F- and t-tests.

A few more things on the model selection, take DF1 as an example:

$$\pi_{t+12}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF1)$$

We start with $L_1 = L_2 = 4$, $L_3 = 1$ and $n = 11$ "supply shock" variables, see the data section for which 11 variables and no dummy variables. Then we stepwise reduce the model further by removing the coefficient with smallest absolute t-statistic, this continues until all coefficients have an absolute t-statistic of 2 or greater, the choice of 2 is quite arbitrary but was chosen by considering that we didn't want to remove a variable which could have importance but not strongly statistically significant but still able to reduce the number of coefficients to about 5-12. Then, we use an F-test to see whether the seasonal dummy variables are jointly statistically significant when added to the model, if they are we add them to the model.

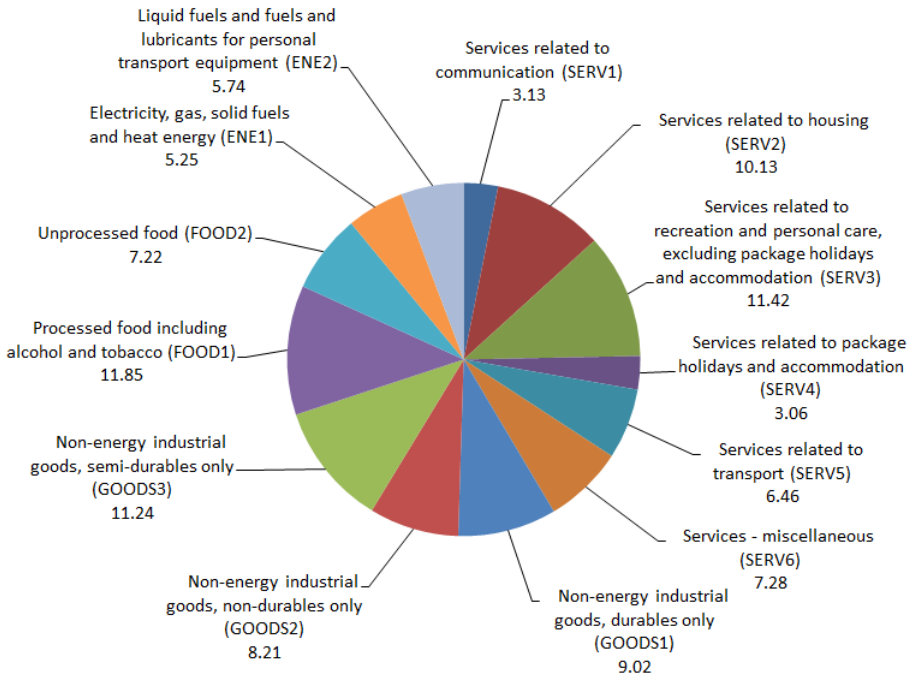
For the iterated forecasts we use a subsidiary AR model for unemployment as Gordon (1998) did. The lag-length for L_1 and L_2 is decided by AIC, and if the seasonal dummy variables should be included by an F-test.

6 Data

For HICP data, the sample covers monthly data from 1996m1 to 2012m11 for Euro Area as a whole, all Euro Area countries and the sub-components for both Euro Area and the different Euro Area countries. The weight series are annual and covers the same sample period. Figure 6.1 show the different sub-components and how they add up to HICP, the weights are for the Euro Area as whole, different countries have of course different weights depending on the size of the economy.

Figure 6.1: Euro Area HICP sub-components

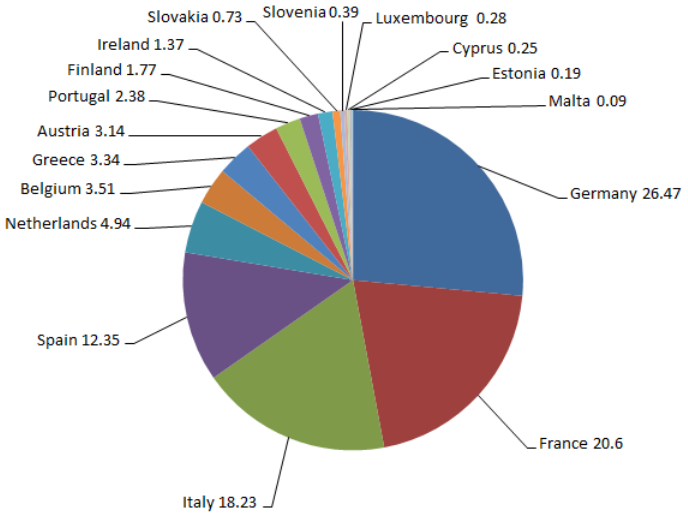
The table below shows the different sub-components we have chosen. The number is the sub-component's weight for 2012, the weights add up to 100.01 due to rounding errors. We will refer to the sub-component by the name in the parenthesis. You may visit Eurostat's Ramon database to see which product groups are included in the sub-component.



Data: Eurostat

The Euro Area today consist of 17 European Union (EU) member states which all have adopted the euro. From 1996-2000, 11 countries aggregates up to the Euro Area HICP, Greece joined 2001, Slovenia in 2007, Cyprus and Malta in 2008, Slovakia in 2009 and lastly Estonia in 2011. Figure 6.2 show the different country weights for 2012.

Figure 6.2: Euro Area HICP country weights
The graph below shows the Euro Area countries and their weight for Euro Area HICP in 2012, the weights add up to 100.03 due to rounding errors.



Data: Eurostat

All inflation data is non-seasonally adjusted and the source is Eurostat. An important date to remember is 1st of January 1999, the official launch of the euro and when the ECB started setting the interest rate. See figure 6.3 for the historical graph which also suggests a break in 1999m1.

Figure 6.3: Euro area HICP

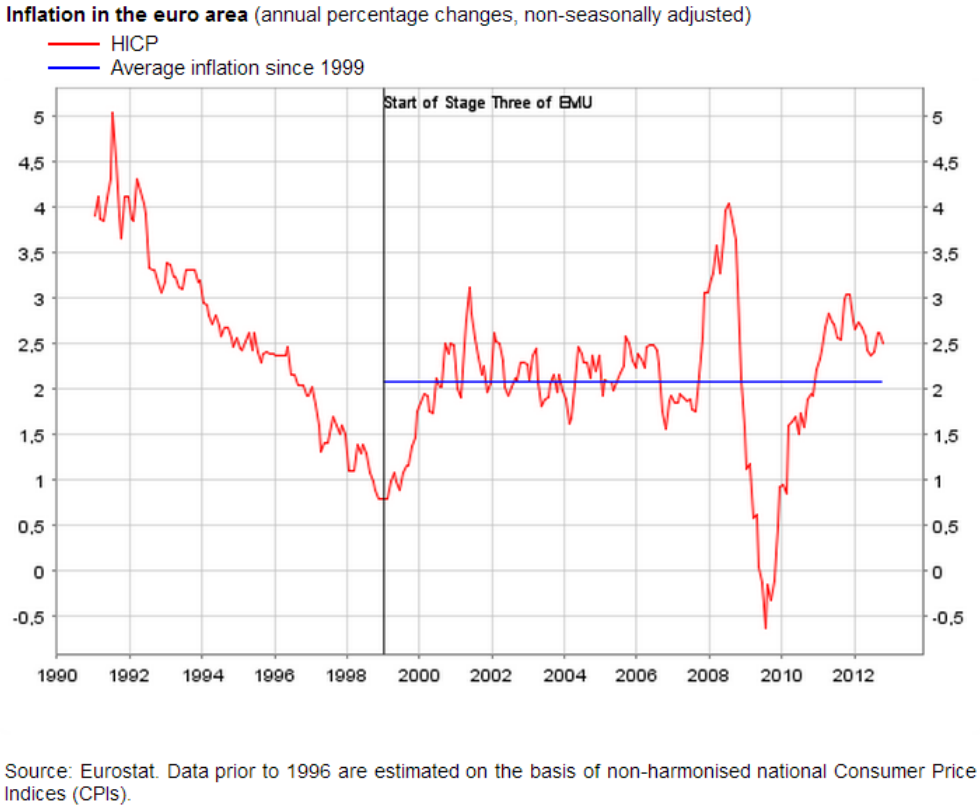
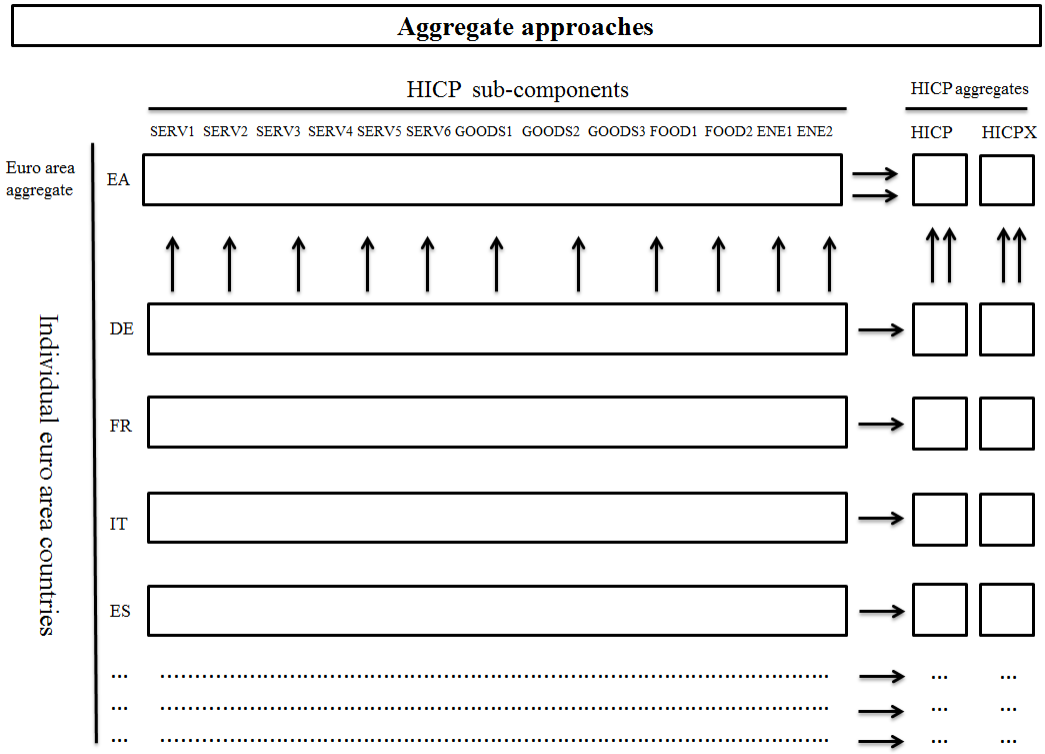


Figure 6.4 shows how one can aggregate the sub-components into for example a country aggregate and then in turn aggregate the country aggregates into a the Euro Area aggregate. However, notice that the result will only be approximate. Figure 6.5 shows how small the difference is.

Figure 6.4: Aggregate approaches

The below figure demonstrates how one can aggregate from the sub-components to Euro Area HICP/HICPX. For example one can aggregate all 17 Euro Area countries SERV1 to get the Euro Area SERV1. We can do this for the rest of the sub-components so we have aggregated all 13 sub-components of Euro Area by using the individual countries sub-components. We can then use these aggregated sub-components to aggregate to HICP/HICPX. Another route would be to get each country's HICP by aggregating it's 13 sub-components and then get Euro Area HICP by aggregating those aggregates.



Source: Benalal et al. (2004) with adjustments

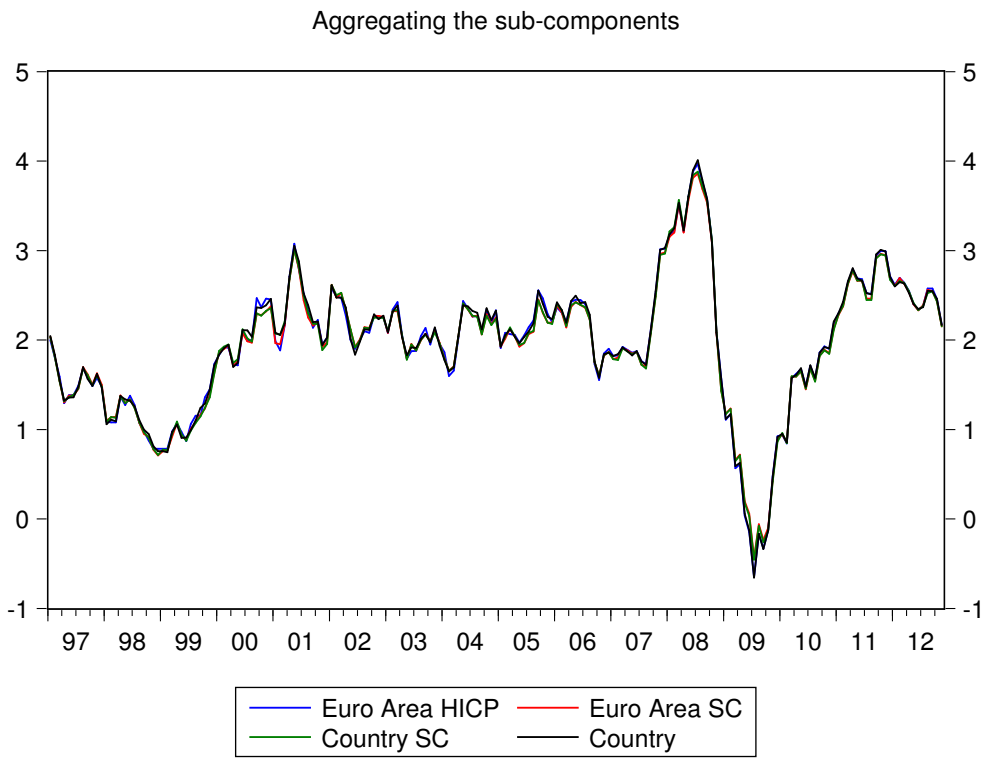
Figure 6.5: Aggregating the sub-components

Euro Area HICP is the year over year change of the official Euro Area HICP Index.

Euro Area SC is when we aggregate from the 13 Euro Area sub-components.

Country SC is when we first aggregate each country's 13 sub-components to that country's HICP and then aggregate all those aggregates to get Euro Area HICP.

Country is when we aggregate all 17 Euro Area countries HICP to get Euro Area HICP, i.e. the second step in Country SC, but here we go directly from the country's HICP and not through the sub-components of the country.



Data: Eurostat

Other than inflation data and weights we also have other variables which may contain information for future inflation, the z variables in our models.

External cost factors

- OIL: Crude oil price in EUR, Hamburgisches WeltWirtschaftsInstitut (HWWI Index), converted from daily data by taking the average index level, then taking Y-o-Y changes
- FOODP: Food prices in EUR, HWWI Index, converted from daily data by taking the average index level, then taking Y-o-Y changes
- MAT: Industrial raw materials in EUR, HWWI Index, converted from daily data by taking the average index level, then taking Y-o-Y changes
- NEER: Nominal effective exchange rate for each country, converted from daily data by taking the average index level, then taking Y-o-Y changes
- ULC: Nominal Unit Labor Cost, Eurostat, converted from quarterly to monthly through linear interpolation and then taking Y-o-Y changes

Activity variables

- GDP: Real GDP, Eurostat, non-seasonally adjusted, converted from quarterly to monthly through linear interpolation and then taking Y-o-Y changes
- UNEMP: Unemployment rate, Ages 15-74, Eurostat, seasonally adjusted
- EMPL: Employment rate, Ages 15-64, Eurostat, non-seasonally adjusted, converted from quarterly to monthly and then seasonally adjusted using X-12 ARIMA
- GYIELD2Y: 2Y Government Benchmark Yield, Macrobond, converted from daily data by taking the average
- GYIELD5Y: 5Y Government Benchmark Yield, Macrobond, converted from daily data by taking the average
- GYIELD10Y: 10Y Government Benchmark Yield, Macrobond, converted from daily data by taking the average

Missing and shorter data series: Note that for some of the small Euro Area countries, the above series are missing or only contain a few observations, if we have fewer than 150 observations for the explanatory variable we throw it away. However, for the unemployment rate, we simply write back the series with the latest value to fill up the history, this shouldn't affect our study much as it was only needed for 3 of the smallest countries and for only about 2 years' worth of data in the estimation window, so it only leads to worse estimation and not cheating out-of-sample forecasts. Also a few of these countries have shorter inflation series for their sub-components, this is especially only for Slovenia and Estonia, for these two countries we moved the estimation window and forecast evaluation period 3 years later.

7 Direct Euro Area HICP Forecast

In this section, we will demonstrate our different forecast models and procedures applied on the Euro Area HICP, both headline (HICP) and core (HICPX). The aim of this section is to show some of the practical difficulties when working with ARMA models and Phillips curve based models, and how we develop our algorithms of model selection to account for some of those.

7.1 ARMA

Let us first view the data, figure 7.1 plots both HICP and HICPX. HICP has averaged 1.95 during this period, a great achievement for the ECB target to maintain inflation rates below, but close to, 2% over the medium term. However, we can see that in 2008 there was first a pickup in inflation which then was followed by a period of low inflation rate, even some deflation in 2009, this is of course related to the financial crisis. HICPX is fairly stable, however interestingly it has been below 2% for the whole period except for two years starting in the end of 2001. HICPX with an average of 1.53 has either failed to work as a measure for the long-run trend in inflation or reflects just the fact that during this period energy prices has grown at a higher rate than 2% per annum.

Figure 7.1: Euro Area HICP

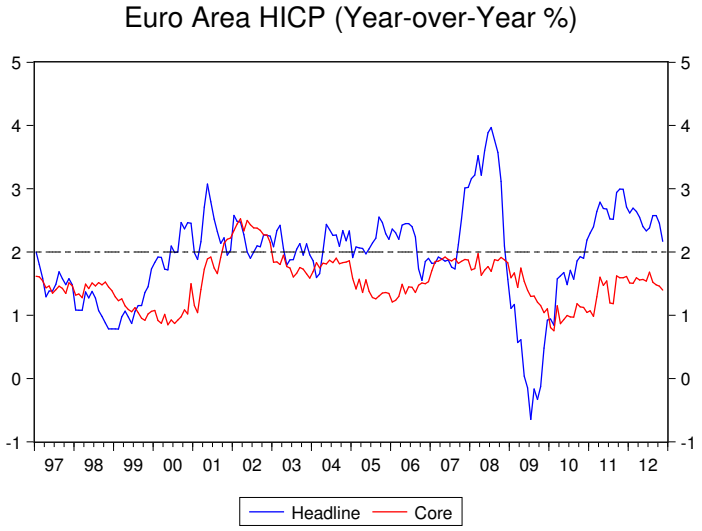
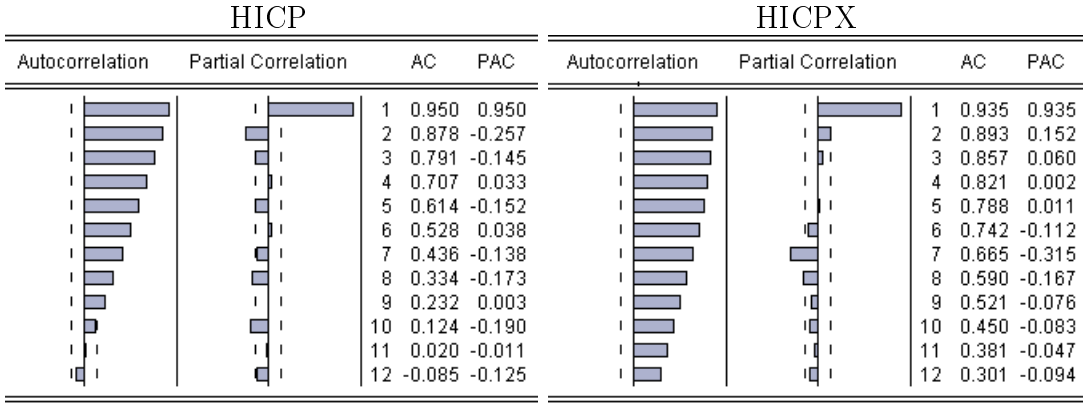


Figure 7.2 shows the autocorrelation and partial autocorrelation for the Y-o-Y series. Both are very similar and show a high and persistent autocorrelation suggesting an AR process and perhaps even a unit-root. This is not surprising as we have Y-o-Y series and monthly observations, so only 2 underlying values out of 12 differs from subsequent observations. The PAC suggests that perhaps an AR(2) or even an AR(1) would most likely be a good fit for both HICP and HICPX.

Figure 7.2: Correlograms



Before we start estimating ARMA models, we want to test if the series contains a unit-root or not. We apply the following ADF test:

$$\triangle \pi_t^{12} = \alpha + \gamma \pi_{t-1}^{12} + \sum_{i=1}^{k} d_i \triangle \pi_{t-i}^{12} + \varepsilon_t,$$

where $\triangle$ is the difference operator, π_t^{12} is the Y-o-Y inflation rate, α is a constant, k is the number of lagged differences and ε_t is the error-term. A time trend is not included as the data shows a stable (non-accelerating) inflation rate and no long-run time trend. We allow for a maximum of both 6 and 16 lagged differences and the model is selected using AIC. Under the null-hypothesis, $H_0 : \gamma = 0$, we have a unit-root and therefore non-stationary series which the usual inference will not be valid for. The alternative hypothesis is $H_1 : \gamma < 0$.

For HICP we cannot reject the null hypothesis of a unit-root when using a constant and 12 lagged difference chosen by AIC (from a maximum of 16 lagged differences). The reported p-value is 0.063, however when only allowing a maximum of 6 lagged differences AIC selects 4 lagged differences and manages to

reject the null of a unit-root with p-value of 0.013. The coefficientγ is -0.08 in both cases, suggesting that the extra lagged differences included in the first model reduced the power of the ADF test making us unable to reject the null-hypothesis. For HICPX the story is similar but the other way around, with 13 lagged differences we reject the null-hypothesis with p-value 0.03 but when only including 6 lagged differences we get a γ much closer to zero and therefore cannot reject the null-hypothesis with a p-value of 0.06.

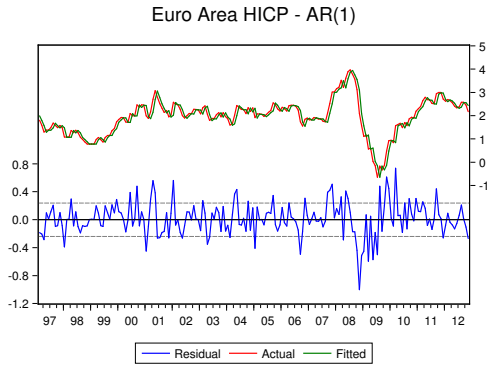
The ADF tests are pretty inconclusive, looking at figure 7.1 again it looks like HICPX show signs of periodic time trends which complicates this even further. One could imagine that we see a falling trend in HICPX through 97-00 and then a rising trend from 01-02, flat although an downtick in 04 and uptick in 06, falling from 08-09 and then rising again from 10 onwards. If this was the case we would need to model in several structural breaks. However, we will continue with two cases, in the first case we assume that we have no unit-root so that the series are stationary and regular statistical inference applies. In the second case we assume that we have a unit-root so we must first difference the series before we start fitting an ARMA model.

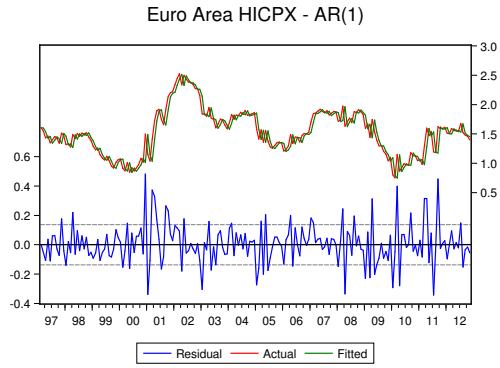
We wish to simulate the real-time forecaster experience as close as possible as well as finding the model which performs best in forecasting. The way we choose to do this is to estimate the same ARMA model recursively using a growing window, and then selecting the model with lowest RMSE. More specifically our first estimation period is 1997m1-2002m1 ($5 * 12 + 1 = 61$) observation of Y-o-Y data, we estimate our ARMA model and produce a 12-step ahead forecast (i.e. we use our forecast values, and not the actual values), so we get a forecast on the 2003m1 observation. We then grow the window with one observation, estimation period 1997m1-2002m2, get new parameter estimates for the same ARMA model and forecast the 2003m2 observation and so on. We then have an out-of-sample forecast series which we can calculate the RMSE for, the RMSE is calculated using the period 2003m1-2012m11 (i.e. all data where we have both the forecast series and actual data). We do this for 91 models (ARMA(p,q), where p = 0-12 and q = 0-6) and select the ARMA model with the lowest RMSE. One reason to estimate the same ARMA model for each window is so that we can plot the parameters of that model to see if its parameters are stable. An alternative is also to use

the model with best in-sample fit for each window, previous research has shown that the best in-sample fit is not necessarily the best for out-of-sample forecasts. However, for completeness we will also report the forecast series produced by trying all 91 models for each window and selecting the best in-sample fit model using an AIC. Note that this series may be a bit irregular as each subsequent observation may have used a different ARMA model. Also note that the AIC is more true to the real-time forecaster experience as both model selection and model estimation is done using only the data available at the time of the forecast, while in the other case the model selection has used the full sample.

The last step in the Box-Jenkins methodology is model diagnostic, in particular to test if the residuals behave as white-noise. This is done with the Ljung-Box test with 12 autocorrelations included and a p-value equal or greater than 0.05 is required to pass the diagnostic check. For convenience, we reduce the set of ARMA models we try out to those which cannot reject the null hypothesis of zero autocorrelation up to 12 lags in the whole sample. Because the data is obviously autocorrelated, there must at least be an AR(1) parameter, figure 7.3 shows the fitted model and it's residuals as well as the correlogram of the residuals. The good fit is due to our use of Y-o-Y data and should be taken with caution. From the correlogram we see a strong autocorrelation to the 12th lag, adding a MA(12) parameter fixes this so in all our ARMA models we will also have included a MA(12) parameter. However, we do a t-test to see that the MA(12) parameter is statistically significant, if the absolute value of the t-statistic is less than 1.5, we remove it. So our first set of ARMA models will be ARMA(p,q+12), where p=1-12 and q=0-6, with MA(12) included if significant. And several of these models won't be reported as they won't fulfill the diagnostic check. It's important to note that there is a possibility that during some windows the residuals of the model could be autocorrelated as we only require that the residuals of the whole sample fit is not autocorrelated. However, in our AIC approach, we require that each model in each window must produce white-noise for residuals, a problem with this is perhaps in one window, none of our models satisfies this requirement but we deem that highly unlikely.

Figure 7.3: Euro Area - AR(1)





HICP

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.240	0.240	11.163	0.001
		2	0.163	0.111	16.302	0.000
		3	-0.004	-0.070	16.305	0.001
		4	0.105	0.110	18.448	0.001
		5	-0.046	-0.088	18.866	0.002
		6	0.075	0.082	19.976	0.003
		7	0.097	0.099	21.843	0.003
		8	0.007	-0.084	21.853	0.005
		9	0.072	0.099	22.906	0.006
		10	-0.039	-0.091	23.222	0.010
		11	0.017	0.015	23.282	0.016
		12	-0.434	-0.454	61.875	0.000

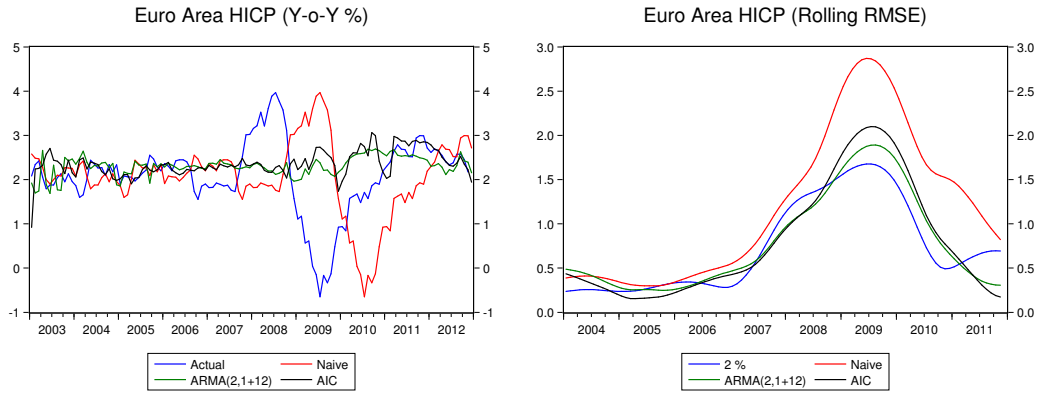
HICPX

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.141	-0.141	3.8557	0.050
		2	-0.013	-0.034	3.8910	0.143
		3	0.035	0.029	4.1307	0.248
		4	0.011	0.020	4.1553	0.385
		5	0.120	0.129	6.9824	0.222
		6	0.256	0.303	19.955	0.003
		7	0.005	0.116	19.961	0.006
		8	-0.021	0.011	20.046	0.010
		9	0.029	0.005	20.210	0.017
		10	0.000	-0.041	20.210	0.027
		11	0.102	0.012	22.351	0.022
		12	-0.153	-0.264	27.160	0.007

Case1:

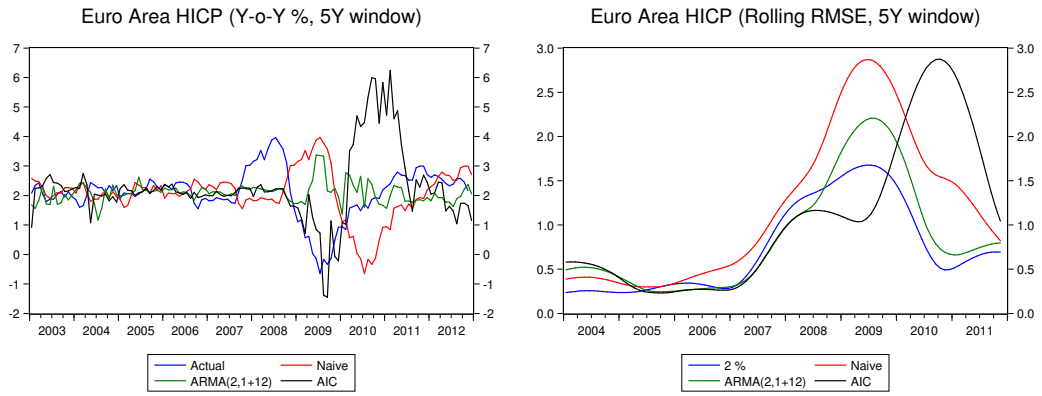
Figure 7.4 shows the results from the above exercise, the model selected was ARMA(2,1+12), which is nice as it's also parsimonious. The relative RMSE is 0.66, which is a huge improvement from the naive forecast. The AIC had a relative RMSE of 0.70 which is slightly worse but still very good. We hypothesized that due to the sharp increase in inflation during 2008 and the following drop in 2009 would severely hurt the naive forecast, and all other forecasts based on lagged dependent variables because exactly when the model starts picking up the increase in inflation, actual inflation falls sharply. However, looking at the rolling RMSE in figure 7.4 we see that it's not only during 2008-2010 that the naive forecast underperforms ARMA(2,1+12), ARMA(2,1+12) outperforms if only just slightly the naive forecast almost throughout the whole sample. What's also interesting, suggesting that the inflation rate one-year ahead will always land on ECB's goal of 2% outperforms even our ARMA(2,1+12) model, with a relative RMSE of 0.63, however it may have been hard to a priori guess that the ECB would be so successful at targeting the inflation rate. The conclusion we draw from this exercise is that during the rather calm period of 2001-2007, the inflation rate was very close to the target of 2% and most univariate models would do quite good at predicting the inflation rate, just as the naive forecast, however due to the strong swings in inflation rate brought upon the financial crisis these univariate models could neither predict nor adjust fast enough.

Figure 7.4: Euro Area HICP - Forecasts & Rolling RMSE



Perhaps we could get better forecasts by using a moving window instead of a growing window as we clearly see two different periods, the first period characterized by a stable inflation rate around 2% and the second period of a much more dramatic swing in the inflation rate. Figure 7.5 shows the same exercise as above but with a moving window of 5 years, now a ARMA(2,4+12) model was selected, however the relative RMSE is worse at 0.73. The performance of AIC is even worse at a relative RMSE of 0.92, trying to forecast inflation in the middle of the financial crisis, putting more weight on recent data (shorter estimation sample), results in erratic behavior of the AIC due to model shift. What is however interesting with the AIC, is that it was somewhat able to predict the fall in inflation rate in 08-09 but then overshoots significantly afterwards when the inflation rate rises. Other window sizes give similar results.

Figure 7.5: Euro Area HICP - 5 year sample window



In figure 7.6 we show the rolling estimates of the parameters in the ARMA(2,1+12) both using the growing window and the moving window. We have two windows when using a moving window that made the estimated model unstable, got non-invertible MA-roots. A 5-year window is perhaps a bit too short. From the left graph in figure 7.6 we can clearly see that the parameters are stable from 2003 to 2010 and then the parameters start drifting a little. However, we conclude that the parameter estimates are stable. What most likely drives the results is the constant's estimate of near 2 most of the time.

Figure 7.6: Euro Area HICP - Rolling Coefficients Estimates

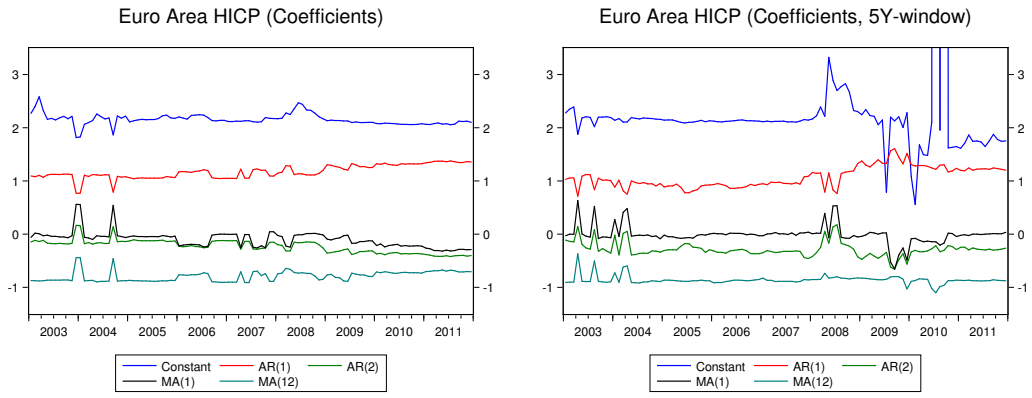
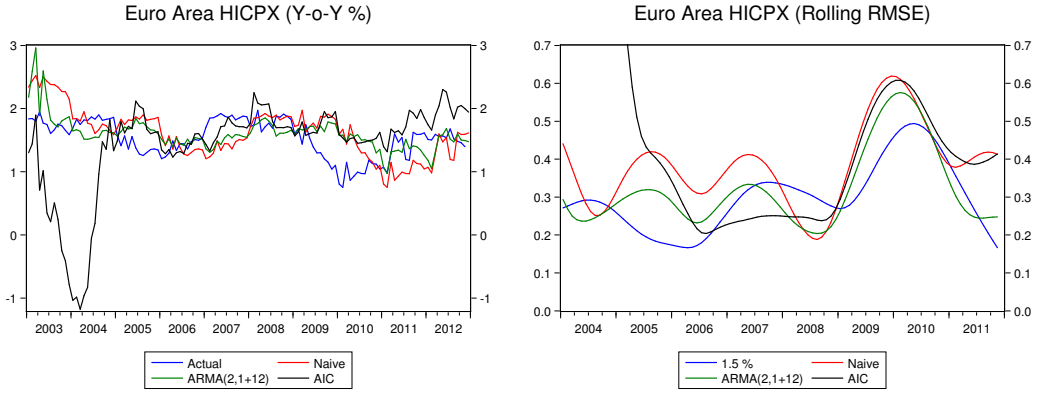


Figure 7.7 shows the first exercise but now with HICPX instead of HICP. The AIC methodology performs awful before it gets a large enough sample. Now an ARMA(1,6+12) got chosen with a relative RMSE of 0.83, which is a significantly gain from the naive forecast. However, it seems the issue is the same as when we worked with HICP, a constant of the periods average HICPX (about 1.5%) outperforms all the other univariate models, relative RMSE of 0.80. But we want to yet again stress the point that it's not easy to know this constant beforehand.

Figure 7.7: Euro Area HICPX

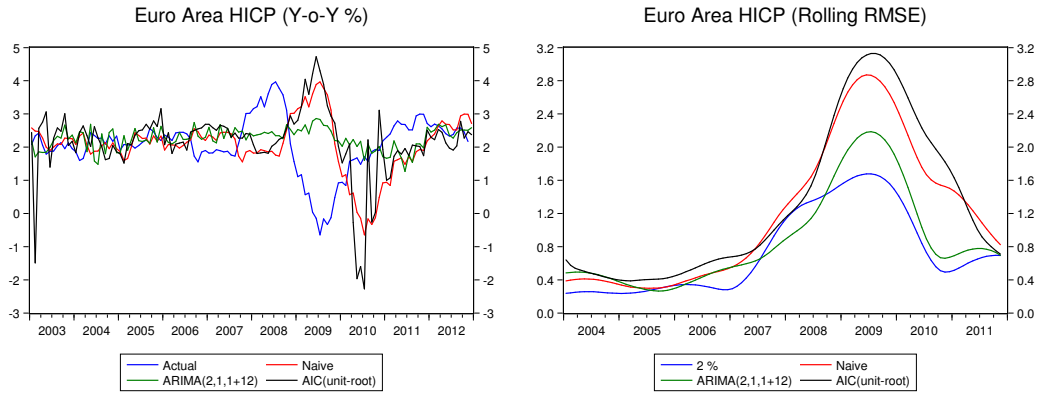


Case 2:

Now we assume that the Y-o-Y data has a unit-root so we estimate ARIMA models instead of ARMA, the approach is exactly the same as earlier except that we difference the dependent variable.

Figure 7.8 shows the results, an ARIMA(2,1,2+12) model was chosen with relative RMSE of 0.90, so still better than the naive approach but not nearly as good as our ARMA(2,1+12) model. We get similar results for HICPX as well. So we conclude that it's much better to model using the direct Y-o-Y series than the difference Y-o-Y series when using an ARMA model. The initial problem why we may wanted to difference the series due to a possible unit-root may not be so serious; as we managed to get good out-of-sample forecasts and also our unit-root tests may have lacked the power to reject the null hypothesis of a unit-root. For example Culver and Papell (1997) show that while it's hard to reject the null hypothesis of a unit-root for individual countries inflation rate, taking the cross-section into account makes the test significantly stronger. In the literature you will find cases where they have modeled inflation as a unit-root process and in others where they have not. Therefore, in our Phillips curve based models, we have both variants so we can compare to both literature and to see which gives the best forecast performance.

Figure 7.8: Euro Area HICP - Forecasts & Rolling RMSE



7.2 Phillips curve based models

In this subsection we will demonstrate our use of both a direct and iterated Phillips curve based model. Hopefully some of the issues involving estimation and the out-of-sample properties become clear.

The two specifications we will use our DF1 and IF1, the discussion generalizes to the other models quite straightforward as only some variable transformations are needed.

$$\pi_{t+12}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t-i} + \sum_{i=1}^{n} \sum_{j=0}^{L_{3,i}} \delta_i z_{i,t-j} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+12} \qquad (DF1)$$

$$\pi_{t+1}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF1)$$

Let us start with the DF1. As in the previous subsection we start our first estimation period 1997m1-2002m1, however some of the "supply shock" variables don't start before 1998 (for example unit labor cost index starts at 1997 so our Y-o-Y series starts at 1998) so we adjust the sample period to 1998m5-2002m1, meaning we have only 45 observations, if we were to start with $L_1 = L_2 = 4$, $L_3 = 1$, $n = 11$ and 11 seasonal dummy variables. We would have $1 + 5 + 5 + 11 * 2 + 11 = 44$ parameters to estimate, i.e. almost as many parameters as observation making the model over-fit and sure to result in bad out-of-sample forecasts. Therefore we never start with the 11 seasonal dummy variables, so we start with estimating 33 parameters and reduce by removing the parameter with smallest absolute t-statistic and re-estimating the reduced model, we continue this until all parameters has a absolute t-statistic of 2 or greater. Then hopefully we have a parsimonious model and can check if all 11 seasonal dummy variables are jointly significant if added to the reduced model, if they are we add them of course. When we have arrived at our reduced model we easily get π_{t+12}^{12} (2003m1 Y-o-Y inflation rate) by fitting the model, we then grow the window with one observation and redo the model reduction and continue till we have produced the full out-of-sample forecast series.

Figure 7.9 show the results, the performance of the DF1 is miserable. It

performs in line with the naive forecast from 2004-2008 but it's unable to capture the dynamics during the financial crisis, in fact this period makes the model unstable. The reason for this is the instability of the parameters, which is among other things a symptom of the model selection procedure which can be seen as data mining. To clarify this further, see figure 7.10 that estimate the same equation, the one chosen by our procedure for 2011m6, but one uses the sample 1998m5-2011m6 and the other 1998m5-2012m11, only 17 more observation is added to get the full sample. It's is clear from the red-line that we managed to find a good fit for the period 1998m5-2011m6 but that using those estimates for out-of-sample forecast performs very badly. Also by using the same parameters but estimated with the full sample results in a worse in-sample fit, 0.28 vs. 0.74 for the shorter sample, as the parameters were not selected by our model reduction procedure. However, we also find that some parameters that were very significant no longer are when re-estimating the same equation, but now with 17 more observations, see table 7.1. Figure 7.11 show the parameter estimation both for a growing window and a rolling fix window of 5 years. Interestingly the parameters look rather stable before the financial crisis and then they change dramatically and afterwards seem to revert back slightly. The issue doesn't seem that we have too many parameters in our reduced model as we only have 8 in this case but because we reduce our model from a large set of parameters we will almost always be able to find those that give a good in-sample-fit but as we saw in this exercise, a great in-sample fit doesn't necessarily lead to a good out-of-sample forecast, especially considering the dramatic shifts in 2007-2010. This is in agreement with Stock and Watson (1999), which found that "supply shock" variables are statistically significant in full-sample specifications but produce worse out-of-sample forecasts when included.

Figure 7.9: Euro Area HICP - Forecasts & Rolling RMSE

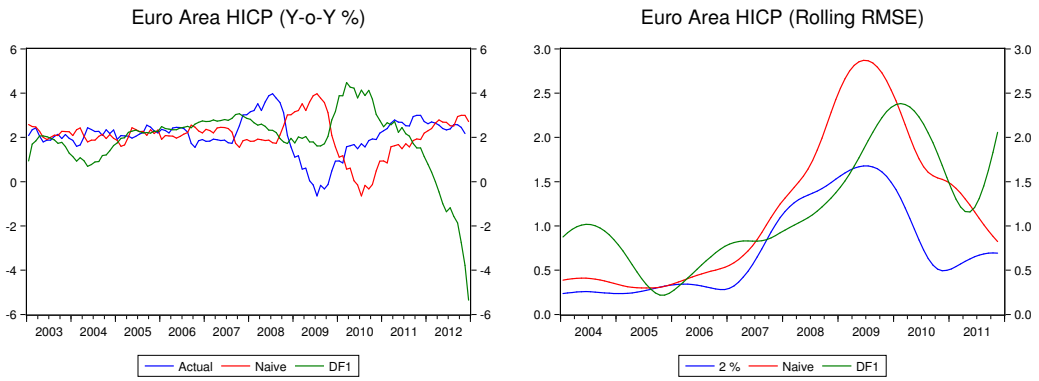


Figure 7.10: Euro Area HICP - Forecasts

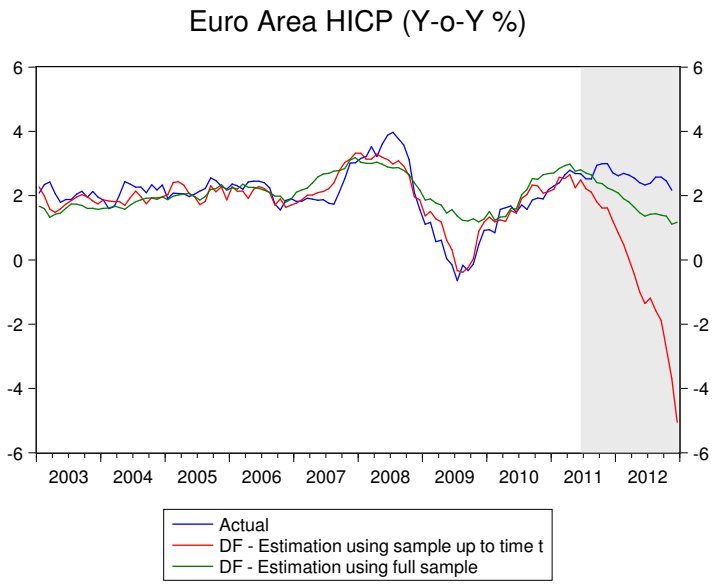
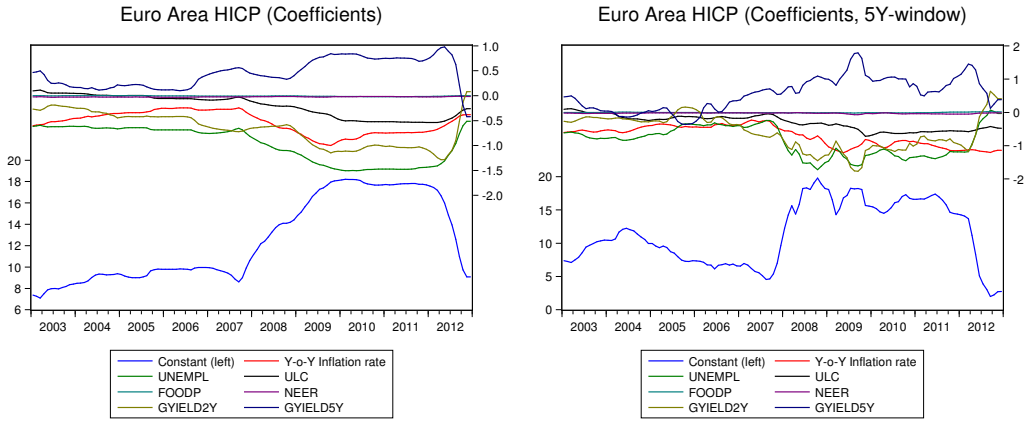


Table 7.1: Coefficient instability, full sample estimate in parenthesis

Variable	Coefficient	t-statistic
α	17.8 (9.08)	22 (9.5)
π_t^{12}	-0.74 (-0.38)	-11 (-3.7)
u_t	-1.48 (-0.52)	-19 (-6.4)
ULC_t	-0.53 (-0.26)	-17 (-6.2)
$FOODP_{t-1}$	-0.02 (0.00)	-6.6 (0.1)
$NEER_t$	-0.02 (-0.02)	-3.6 (0.6)
$GYIELD2Y_{t-1}$	-1.04 (0.08)	-7.3 (0.59)
$GYIELD5Y_{t-1}$	0.75 (-0.42)	4.7 (-2.4)

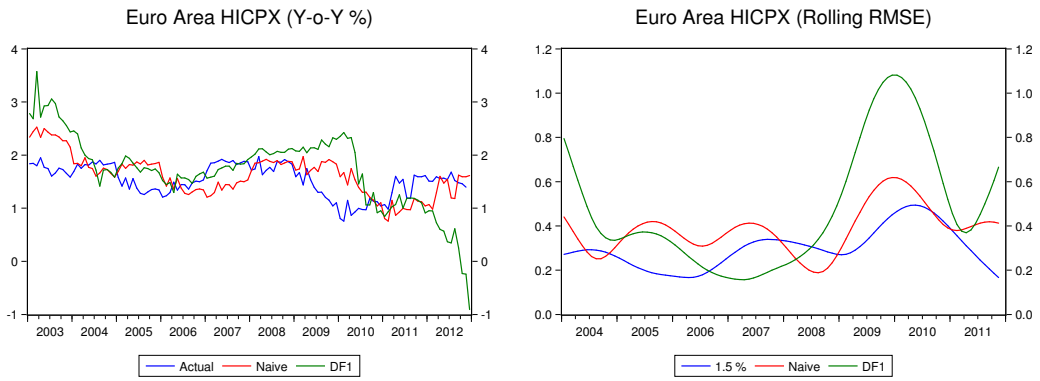
Figure 7.11: Euro Area HICP - Rolling Coefficients Estimates



We noted an issue with including GDP, even though GDP can help predict inflation in our exercise, including it makes us to break the true real-time forecasting experience as the release of GDP numbers is lagged up to several months, usually 3 or more months. We could of course use a lag of GDP to make it more of a real-time forecasting experience but some first evidence suggested it wouldn't be significant, so we decided not to include GDP as a potential variable.

For HICPX the results are similar, good performance before the financial crisis but terrible under, however we return to quite good performance directly after the financial crisis, see figure 7.12.

Figure 7.12: Euro Area HICPX - Forecasts & Rolling RMSE



Now let us turn to IF1:

$$\pi_{t+1}^{12} = \alpha + \sum_{i=0}^{L_1} \beta_i \pi_{t-i}^{12} + \sum_{i=0}^{L_2} \gamma_i u_{t+1-i} + \sum_{i=1}^{11} \eta_i D_i + \varepsilon_{t+1} \qquad (IF1)$$

There is only two difference regarding implementation from DF1. First that we now must make an subsidiary AR forecast of the unemployment rate, this is simply done by testing 12 AR models, with lags 1-12 and select the model with best AIC for the active window. The second difference is that instead of using the t-statistic to reduce the model we select the model with the lowest AIC from the 25 models ($L_1 = L_2 = 4$, so $5 * 5 = 25$) that we run. And then like we did for the DF1, add the seasonal dummies if they are jointly significant (with a p-value of 0.05 or lower) by the F-test.

See figure 7.13 and 7.14 for HICP and HICPX respectively. We see that this model works very well, the relative RMSE for HICP is 0.70 and for HICPX is 0.69. Remember the mean for the respective period gave an relative RMSE of 0.66 for HICP and 0.83 for HICPX, meaning we beat it for HICPX and successfully manage to model some of the dynamics. IF1 worked so well so we became slightly worried that we were using data which wouldn't have been available at time t when producing our subsidiary forecast of unemployment. In some countries Eurostat's release date for the unemployment rate for the same month as the inflation rate lags, however in others like Finland, unemployment data is actually released a couple days before. However, we tested using only data for the unemployment

rate up to time $t-1$ and it only affected the performance slightly, it reduces the relative RMSE for both HICP and HICPX by only 0.01 units. So we feel comfortable that we aren't tricking ourselves.

Figure 7.13: Euro Area HICP - Forecasts & Rolling RMSE

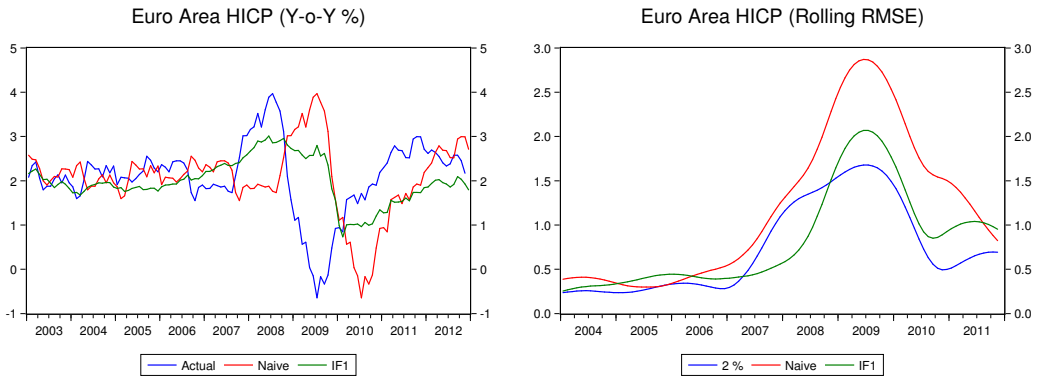
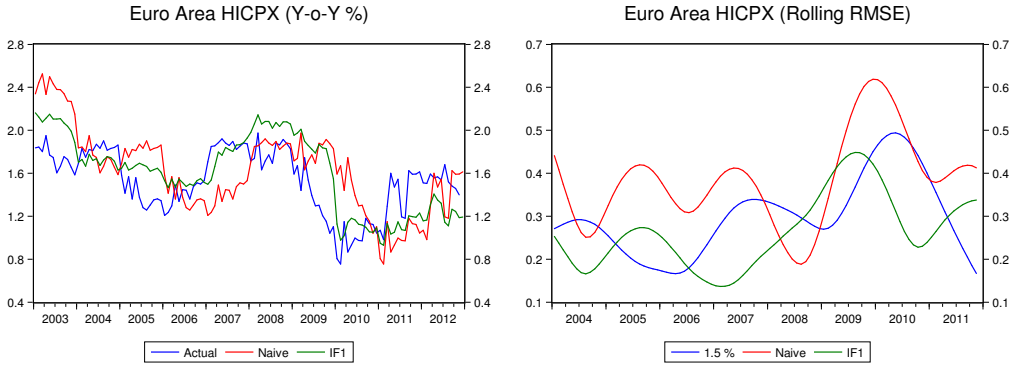


Figure 7.14: Euro Area HICPX - Forecasts & Rolling RMSE



8 Indirect Euro Area HICP Forecast

In this section, we apply our models to forecast the sub-components of Euro Area HICP and to see if aggregating these sub-components outperforms the direct forecasts. In subsection 8.1 we work with the 13 sub-components, which we can aggregate up to both HICP/HICPX. In subsection 8.2 we look at making HICP/HICPX forecasts at the country level and then aggregating up to Euro Area. In subsection 8.3 we compare the different aggregating routes and perform an optimal aggregate.

8.1 Forecasts through HICP Sub-components

Table 8.1 show the performance of our different models on forecasting the different sub-components of Euro Area HICP. IF1 and IF2 outperform all models by far. The DF models are at a huge disadvantage as most of them end being unstable after the financial crisis as the big change in the environment strongly affected the parameter estimates. The ARMA-models do generally quite well in comparison to the naive forecast (AO) that the Y-o-Y inflation rate one-year ahead will be the same as it's today. AIC does very well for HICP but terrible for HICPX and in general doesn't seem to improve on AO. You can see from the AO column that energy prices is by far the most volatile sub-component and we gain huge forecast improvements by using an IF1 or IF2 model on these sub-components.

Note also that we gain in forecast accuracy by aggregating the sub-components to HICP/HICPX, 0.49 and 0.58 than the direct forecast of 0.66 and 0.67.

Table 8.1: Forecast performance of Euro Area HICP/HICPX by going through sub-components

Here we make forecasts of Euro Area 13 HICP sub-components and then aggregate them up to HICP and HICPX, called HICP Agg and HICPX Agg respectively. The value under the column AO is in RMSE (percentage points) and under all other columns it's relative RMSE, to get their RMSE in percentage points you can simply just multiply with the value under AO. The forecast evaluation period which the RMSE is calculated from is 2003m1-2012m11.

Sub-component	AO	ARMA	AIC	DF1	DF2	DF3	DF4	IF1	IF2	IF3	IF4
HICP Agg	1.31	0.66	0.91	1.24	1.17	1.51	1.45	0.65	**0.49**	1.05	1.08
HICPX Agg	0.41	0.90	1.12	1.70	1.61	2.18	1.94	0.69	**0.58**	1.26	1.26
HICP	1.31	**0.66**	0.70	1.33	1.16	1.46	1.37	0.70	**0.66**	1.07	1.14
HICPX	0.41	0.83	2.05	1.66	1.58	1.92	1.89	0.69	**0.67**	1.22	1.21
SERV1	1.24	1.00	2.03	2.49	2.20	3.73	3.51	**0.78**	0.80	1.23	1.24
SERV2	0.22	1.19	1.84	2.98	3.22	3.10	2.89	**0.96**	1.13	1.32	1.46
SERV3	0.68	1.17	1.17	1.75	1.49	1.91	1.73	0.63	**0.42**	1.18	1.18
SERV4	2.31	0.80	0.86	0.94	0.92	0.86	0.95	**0.62**	0.72	0.91	0.93
SERV5	0.84	0.72	0.80	1.16	1.03	0.96	1.22	0.59	**0.57**	0.86	0.97
SERV6	1.02	1.17	1.57	1.91	1.79	1.83	1.54	0.68	**0.65**	0.96	1.01
GOODS1	0.76	0.70	0.80	0.85	0.72	1.09	1.02	**0.59**	0.60	0.94	0.94
GOODS2	0.51	0.83	0.96	1.36	1.33	1.86	1.63	**0.64**	0.67	1.01	0.97
GOODS3	0.78	0.80	1.11	1.29	1.29	1.82	1.73	**0.72**	0.73	1.08	1.07
FOOD1	2.31	0.71	0.95	1.12	0.99	1.43	1.28	0.51	**0.41**	1.02	1.09
FOOD2	2.58	0.71	1.42	1.33	1.13	1.68	1.52	0.65	**0.51**	1.21	1.19
ENE1	6.56	0.62	0.74	0.92	0.79	1.28	1.42	0.52	**0.35**	0.88	1.00
ENE2	19.00	0.54	0.80	0.85	0.91	1.47	1.50	**0.53**	0.55	0.94	0.94

If using sub-components help forecast the HICP/HICPX inflation, then one could most likely get better country forecasts by using the country's sub-components to forecast the country's HICP/HICPX first and then aggregate to Euro Area, this is called Euro Area Agg in table 8.2. And indeed we get some forecast improvement but it's so little it's negligible. However, again IF1 and IF2 perform very well, but now IF2 seems to beat IF1 much more often. It's interesting that the other models gain huge improvement when aggregating from country forecasts to Euro Area, see for example DF1 which relative RMSE for HICP/HICPX goes from 1.24(1.70) to 0.79(0.90). Perhaps this is a result from that DF1 is at times unstable and when we aggregate some of the forecast errors mitigate each other making the aggregate a bit more stable.

Table 8.2: Forecast performance of countries HICP/HICPX by going through sub-components

Here we make forecasts for each country's 13 sub-components and then aggregate them up to HICP and HICPX. Euro Area Agg is when we further aggregate the countries aggregated forecasts. The table reports the aggregated forecasts RMSE, the value without parenthesis is for HICP while the value in parenthesis is for HICPX. The value under the column AO is in RMSE (percentage points) and under all other columns it's relative RMSE, to get their RMSE in percentage points you can simply just multiply with the value under AO. The forecast evaluation period which the RMSE is calculated from is 2003m1-2012m11.

Country	AO	ARMA	AIC	DF1	DF2	DF3	DF4	IF1	IF2	IF3	IF4
Euro Area Agg	1.31(0.41)	0.69(0.77)	0.76(0.94)	0.79(0.90)	0.71(1.04)	1.09(1.47)	1.10(1.52)	0.62(0.61)	**0.47**(**0.57**)	0.95(1.06)	0.98(1.09)
Euro Area	1.31(0.41)	0.66(0.90)	0.91(1.12)	1.24(1.70)	1.17(1.61)	1.51(2.18)	1.45(1.94)	0.65(0.69)	**0.49**(**0.58**)	1.05(1.26)	1.08(1.26)
Germany	1.23(0.73)	0.75(0.66)	1.04(1.25)	1.22(0.92)	0.91(0.91)	1.48(1.37)	1.38(1.34)	0.65(0.61)	**0.56**(**0.54**)	0.99(1.09)	1.00(1.13)
France	1.42(0.36)	0.67(1.03)	0.82(1.81)	0.85(1.22)	0.77(1.09)	1.11(1.96)	1.12(1.66)	0.63(0.70)	**0.52**(**0.67**)	0.98(1.31)	1.01(1.32)
Italy	1.25(0.54)	0.72(0.75)	0.77(0.86)	0.89(1.21)	0.78(1.19)	1.10(1.43)	1.05(1.49)	0.66(0.73)	**0.47**(**0.66**)	0.96(1.12)	0.96(1.14)
Spain	1.99(0.82)	0.70(0.99)	0.76(1.17)	0.87(1.46)	0.81(1.34)	1.44(1.78)	1.48(1.80)	0.63(**0.59**)	**0.50**(**0.59**)	1.02(0.91)	1.06(0.95)
Netherlands	1.09(0.84)	0.90(0.92)	1.10(1.07)	1.41(1.34)	1.51(1.20)	1.81(1.83)	1.55(1.58)	0.83(0.70)	**0.74**(**0.78**)	1.24(1.17)	1.20(1.11)
Belgium	2.16(0.64)	0.60(0.71)	0.73(0.84)	0.71(0.71)	0.60(0.72)	0.84(1.28)	1.01(1.26)	0.62(0.54)	**0.42**(**0.44**)	0.81(0.91)	1.00(0.96)
Greece	1.93(1.14)	**0.51**(0.88)	0.73(1.09)	0.88(1.31)	1.05(1.32)	2.07(1.81)	2.28(1.90)	0.56(**0.78**)	0.54(0.92)	1.00(1.04)	0.99(1.04)
Austria	1.45(0.71)	0.67(0.80)	0.75(0.87)	1.06(1.29)	0.85(0.97)	1.09(1.28)	1.09(1.17)	0.65(0.64)	**0.55**(**0.55**)	0.95(0.93)	0.96(0.99)
Portugal	1.82(1.16)	0.78(0.97)	0.85(1.20)	1.17(1.25)	0.98(1.05)	1.30(1.45)	1.22(1.31)	0.69(0.66)	**0.53**(**0.60**)	1.03(1.01)	1.06(1.03)
Finland	1.39(0.67)	1.04(1.43)	1.15(1.59)	1.63(2.48)	1.37(2.08)	1.04(1.45)	1.09(1.31)	0.69(**0.88**)	**0.64**(0.98)	0.94(1.00)	0.98(0.99)
Ireland	2.08(1.60)	1.00(1.10)	1.13(1.24)	1.57(1.89)	1.31(1.72)	1.21(1.33)	1.15(1.29)	0.69(0.82)	**0.60**(**0.70**)	0.95(0.98)	0.95(1.00)
Slovakia	2.94(1.85)	0.74(**0.82**)	1.08(1.59)	1.73(2.09)	1.66(2.14)	1.79(1.38)	1.58(1.38)	0.77(1.02)	**0.64**(0.90)	1.05(1.02)	1.08(1.09)
Slovenia	2.45(1.61)	0.69(1.01)	0.89(1.13)	1.30(1.50)	1.15(1.66)	1.62(1.61)	1.62(1.53)	**0.55**(**0.75**)	0.60(1.01)	1.04(1.25)	1.04(1.25)
Luxembourg	2.18(0.47)	0.63(**0.64**)	0.75(0.94)	1.18(2.57)	0.94(2.06)	1.14(1.56)	1.12(1.22)	0.67(0.77)	**0.54**(0.72)	0.96(1.05)	0.96(1.02)
Cyprus	2.35(1.36)	0.67(0.75)	0.64(0.85)	0.85(0.99)	0.87(1.04)	1.71(1.49)	1.51(1.27)	0.56(0.70)	**0.54**(**0.68**)	0.93(0.99)	0.97(1.05)
Estonia	4.83(2.59)	0.77(0.76)	0.77(0.90)	0.97(1.40)	0.93(1.42)	1.44(1.61)	1.38(1.52)	0.64(0.68)	**0.38**(**0.61**)	1.13(1.23)	1.12(1.23)
Malta	2.31(1.83)	0.63(0.73)	0.79(0.89)	0.69(0.87)	0.64(0.74)	1.14(1.07)	1.24(1.22)	0.61(0.64)	**0.52**(**0.60**)	0.95(0.95)	0.99(0.98)

8.2 Forecast through Country Forecasts

Table 8.3 show how well our models perform in forecasting individual Euro Area countries, again the IF1 and IF2 outperform the rest, however IF2 beats IF1 a few more times. The performance for the different countries are quite similar, the largest forecast errors we get for Estonia, where with using the best model we get an RMSE of 1.88 ($4.83 * 0.39$). Again we see that we can improve on the Euro Area HICP/HICPX forecast by first forecasting all countries HICP/HICPX and aggregate them up to Euro Area.

Table 8.3: Forecast performance of countries HICP/HICPX by going through country forecasts

Here we make forecasts on both a country's HICP and HICPX and then aggregate all countries forecasts up to Euro Area HICP and HICPX, which is Euro Area Agg in the table. The table reports the aggregated forecasts RMSE, the value without parenthesis is for HICP while the value in parenthesis is for HICPX. The value under the column AO is in RMSE (percentage points) and under all other columns it's relative RMSE, to get their RMSE in percentage points you can simply just multiply with the value under AO. The forecast evaluation period which the RMSE is calculated from is 2003m1-2012m11.

Country	AO	ARMA	AIC	DF1	DF2	DF3	DF4	IF1	IF2	IF3	IF4
Euro Area Agg	1.31 (0.41)	0.65 (0.76)	0.72 (0.83)	0.77 (1.01)	0.73 (1.03)	1.16 (1.52)	1.13 (1.53)	0.65 (0.65)	**0.53** (**0.58**)	0.97 (1.05)	1.01 (1.05)
Euro Area	1.31 (0.41)	0.66 (0.83)	0.70 (2.05)	1.33 (1.66)	1.16 (1.58)	1.46 (1.92)	1.37 (1.89)	0.70 (0.69)	**0.66** (**0.67**)	1.07 (1.22)	1.14 (1.21)
Germany	1.23 (0.73)	0.77 (0.67)	0.84 (0.81)	1.04 (1.09)	0.93 (1.01)	1.50 (1.46)	1.37 (1.44)	**0.65** (0.68)	0.69 (**0.67**)	0.97 (1.09)	0.99 (1.09)
France	1.42 (0.36)	0.66 (1.07)	0.76 (1.63)	1.02 (1.23)	0.79 (1.19)	1.13 (1.73)	1.13 (1.64)	0.71 (0.84)	**0.53** (**0.68**)	1.01 (1.34)	1.05 (1.34)
Italy	1.25 (0.54)	0.67 (0.81)	0.74 (0.91)	0.94 (1.14)	0.77 (1.10)	1.10 (1.49)	1.09 (1.41)	0.69 (0.75)	**0.66** (0.79)	0.96 (1.08)	0.98 (1.08)
Spain	1.99 (0.82)	0.69 (1.09)	0.95 (1.44)	0.88 (1.29)	0.92 (1.28)	1.61 (1.78)	1.57 (1.69)	0.62 (**0.66**)	**0.54** (0.69)	1.00 (0.97)	1.07 (0.96)
Netherlands	1.09 (0.84)	1.03 (0.91)	1.19 (0.99)	1.52 (1.75)	1.52 (1.27)	1.83 (1.68)	1.74 (1.70)	0.81 (0.71)	**0.71** (**0.69**)	1.34 (1.17)	1.26 (1.15)
Belgium	2.16 (0.64)	**0.60** (0.61)	0.69 (0.67)	0.66 (0.77)	0.62 (0.84)	1.00 (1.36)	1.04 (1.47)	0.68 (0.48)	0.62 (**0.45**)	1.01 (1.00)	1.01 (0.98)
Greece	1.93 (1.14)	**0.48** (0.99)	0.52 (1.31)	0.78 (1.16)	1.04 (1.51)	2.05 (2.06)	2.03 (2.07)	0.60 (**0.77**)	0.63 (0.98)	0.97 (1.03)	0.96 (1.06)
Austria	1.45 (0.71)	0.72 (0.80)	0.95 (0.96)	1.01 (1.27)	0.91 (1.21)	1.06 (1.51)	1.08 (1.20)	0.73 (0.72)	**0.55** (**0.66**)	1.00 (1.00)	1.05 (1.03)
Portugal	1.82 (1.16)	0.86 (1.18)	0.93 (1.32)	1.22 (1.31)	0.99 (1.13)	1.20 (1.52)	1.19 (1.42)	0.72 (0.70)	**0.56** (**0.69**)	1.03 (1.04)	1.04 (1.05)
Finland	1.39 (0.67)	0.94 (1.12)	0.99 (1.45)	1.62 (2.06)	1.23 (1.98)	1.11 (1.31)	1.04 (1.32)	0.71 (**0.70**)	**0.60** (1.23)	0.97 (0.89)	0.97 (0.89)
Ireland	2.08 (1.60)	0.98 (1.18)	1.15 (1.28)	1.51 (1.82)	1.20 (1.82)	1.41 (1.31)	1.29 (1.22)	0.75 (0.82)	**0.51** (**0.87**)	0.90 (1.10)	0.93 (1.00)
Slovakia	2.94 (1.85)	0.83 (0.94)	0.88 (1.07)	1.71 (1.89)	1.56 (2.23)	1.91 (1.47)	1.45 (1.24)	0.73 (0.87)	**0.63** (**0.78**)	1.03 (0.99)	1.03 (1.01)
Slovenia	2.45 (1.61)	0.85 (1.13)	1.04 (1.45)	1.07 (1.81)	1.02 (1.49)	1.40 (1.51)	1.40 (1.51)	**0.70** (**0.75**)	**0.70** (0.94)	1.01 (1.29)	1.01 (1.25)
Luxembourg	2.18 (0.47)	0.62 (0.71)	0.84 (0.81)	1.26 (2.20)	0.93 (2.23)	1.10 (1.38)	1.10 (1.36)	0.62 (**0.76**)	**0.54** (0.85)	0.99 (1.07)	1.00 (1.06)
Cyprus	2.35 (1.36)	0.54 (0.80)	0.57 (0.91)	0.80 (1.20)	0.79 (0.92)	1.13 (1.38)	1.30 (1.21)	0.55 (**0.70**)	**0.52** (0.72)	0.95 (1.05)	0.98 (1.03)
Estonia	4.83 (2.59)	0.62 (0.78)	0.71 (0.86)	1.06 (1.47)	0.96 (1.32)	1.38 (1.41)	1.38 (1.42)	0.69 (0.66)	**0.39** (**0.54**)	1.15 (1.12)	1.13 (1.16)
Malta	2.31 (1.83)	**0.43** (0.66)	0.46 (0.69)	0.54 (0.82)	0.66 (0.81)	1.14 (1.09)	1.25 (1.17)	0.58 (**0.65**)	0.57 (0.67)	0.95 (0.95)	1.02 (1.00)

8.3 Optimized Euro Area Forecast

In the previous two subsections we aggregated the sub-components in three ways, through Euro Area HICP sub-components, through all countries HICP sub-components and lastly through country HICP/HICPX. Here we will consider two additionally approaches. Starting from all countries HICP sub-components, we can aggregate either to country level and then Euro Area HICP/HICPX, or aggregate to Euro Area HICP sub-components and then to Euro Area HICP/HICPX. The result would be the same, however now, if an intermediate aggregate has a better "direct" forecast, that forecast will be replaced with the "direct" forecast and therefore the two different routes of aggregating will differ. Table 8.4 summarizes the results from the different aggregation approaches. Not very surprising, the optimal forecast do better than the others, however one must take care that in this procedure we break the real-time forecasting experience as the choice of model is based solely on the relative RMSE which we can't know beforehand. What is interesting is that we gain our biggest forecast accuracy improvement by first forecasting the 13-sub components of Euro Area and aggregating up to Euro Area HICP/HICPX. Disaggregating further and forecasting all countries sub-components seem to help for HICP but not at any significant level for HICPX.

Table 8.4: Forecast performance of different aggregate approaches
Here we show the forecast performance from the different ways of aggregating. Euro Area is when we forecast directly with IF2. Euro Area SC is when we first forecast the 13 sub-components of Euro Area with IF2 and then aggregate. Country SC is when we use IF2 to forecast all countries sub-components and then aggregate these up to Euro Area HICP/HICPX. Country is when we use IF2 to forecast each country's HICP/HICPX and aggregate to Euro Area. Optimal SC is when we optimize using the sub-components route. Optimal Country is when we optimize using the country aggregation route. The value under the column AO is in RMSE (percentage points) and under all other columns it's relative RMSE, to get their RMSE in percentage points you can simply just multiply with the value under AO. The forecast evaluation period which the RMSE is calculated from is 2003m1-2012m11.

Sub-component	AO	Euro Area	Euro Area SC	Country SC	Country	Optimal SC	Optimal Country
HICP	1.31	0.66	0.53	0.47	0.53	0.47	**0.46**
HICPX	0.41	0.67	0.58	0.57	0.58	**0.51**	0.53

Figure 8.1 and 8.2 shows the optimal forecasts for HICP/HICPX respectively. We see that we manage to capture a lot of the dynamics and beat the naive forecast by far, and even the average of the period, which we wouldn't be able to know a-priori.

Figure 8.1: Euro Area HICP - Optimal Forecasts

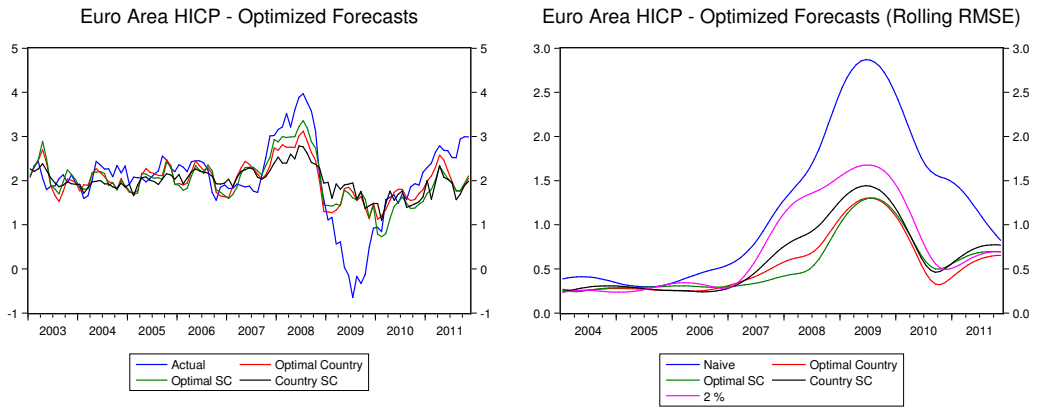
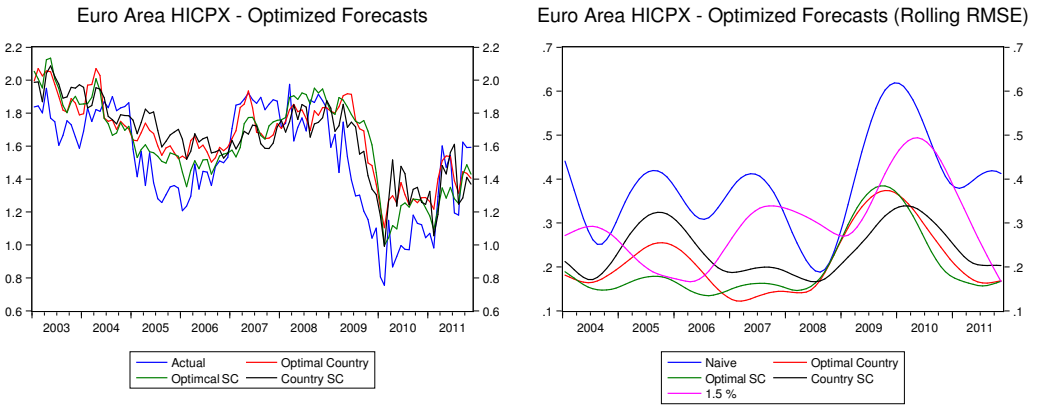


Figure 8.2: Euro Area HICPX - Optimal Forecasts



9 Conclusions

We set out to forecast the Y-o-Y Euro Area HICP/HICPX one-year ahead. We knew that forecasting inflation is notoriously hard, and that even a naive forecast, that inflation over the next year will be the same as it has been during the past year, is hard to beat, or in any case a good AR model. However, we showed that with a simple Phillips curve based model and a basic model selection algorithm that we managed to beat both the naive forecast and our ARMA model during the forecast evaluation period of 2003m1-2012m1. We also showed that we could improve even further on the Euro Area HICP/HICPX by first forecasting their sub-components and then aggregating the forecasts. Most likely the main driver of this result is the huge improvement made in modeling the energy sub-component, which is by far the most volatile component. You can view energy prices as a white noise process, and when trying to forecast Euro Area HICP directly, you will still have unbiased estimates but due to the noise, these estimates will be inexact and lead to worse forecast performance. When we disaggregate into sub-components, we handle energy prices by themselves and in that way filter away the white noise. Also the naive forecast for energy prices is very flawed, say for example that prices rose by 20% the previous year due to a supply shock, then you would most likely not believe that they will rise 20% this year as well, a more realistic assumption in that case would be unchanged prices from today.

We also set out to answer three methodological questions:

Q1. Is it better to use a direct forecast or iterated forecast?

Q2. Should one use past year-over-year (Y-o-Y) or month-over-month (M-o-M) inflation rate as explaining variables?

Q3. Should one model the Y-o-Y inflation rate as a unit-process or not?

A1. In our case it seemed to be clearly better to use a iterated forecast instead of an direct forecast. Perhaps due to our sample where the financial crisis is included but not directly modeled made our parameters too unstable and resulted in a disadvantage to the direct forecast which used more parameters which may also perhaps be more sensitive to larger changes.

A2. It didn't seem to considerable matter if one used Y-o-Y or M-o-M for past inflation. It would have been interesting to see if it mattered if we modeled the

dependent variable as M-o-M instead of Y-o-Y as it did for Hendry and Hubrich (2010) for the US case, they found that using M-o-M instead of Y-o-Y gives better forecast performance.

A3. It is also clear at least from the models that we used that it's better to treat Euro Area inflation as a stationary process and not having a unit-root. This goes against Hendry and Hubrich (2010), but the difference may lie in the difference between Euro Area and US inflation. It seems more common to model inflation as having a unit-root in the US and not for the Euro Area, however it doesn't seem to exist any real consensus in the literature that inflation is non-stationary in the US but stationary in the Euro Area.

Our results confirms with Den Reijer and Vlaar (2003) and Espasa and Albacete (2004) that aggregating the sub-components of Euro Area HICP/HICPX produces better forecasts than to directly forecast Euro Area HICP/HICPX. However, it goes against Hubrich (2005) and Benalal et. al. (2004) which found that using the sub-components only improve the aggregate forecast of HICP for only very short horizon (1-month), however for HICPX they seem to find some improvement by using the sub-components. Our models and the model selection procedure is more similar to theirs than to Den Reijer and Vlaar (2003) and Espasa and Albacete (2004) which uses a VECM, however as we see among our different models, even if they seem similar the performance can vary quite greatly. Also one must consider that we are using a completely different forecast evaluation period, their data set ends at 2002 while our forecast evaluation period starts at 2003. This also means that we regrettably can't more quantitatively compare the results.

Interesting future research would be to apply some of the methodologies covered by Stock and Watson (2006) which gives an extensive overview of forecasting methods using many predictors. They cover forecast combination, dynamic factor models (also called Factor-Augmented Vector Autoregressive Models (FAVAR)), Bayesian model averaging and Empirical Bayes methods. We see a strong potential for some of these models, especially dynamic factor models where one can decompose all the sub-components information into a few factors is intuitively appealing. Furthermore, interesting future research would be to take structural breaks into consideration, we most likely had several structural breaks throughout

our sample, and especially considering that in our sample's time period we have the financial crisis and Euro Area sovereign debt crisis, which led some of our models to be useless.

References

Aron, J. and Muellbauer, J. (2008) New methods for forecasting inflation and its subcomponents: an application to the USA, Working Paper 406, Department of Economics, Oxford University, October 2008

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

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

Box, G. and Jenkins, G. (1976) Time Series Analysis: Forecasting and Control, San Fransisco, Calif, Holden Day

Calvo, G. (1983) Staggered Prices in a Utility-Maximizing Framework.h Journal of Monetary Economics 12(3): 383-98

Camba-Mendez, G. and Kapetanios, G. (2004) Forecasting euro area inflation using dynamic factor measures of underlying inflation, Working Paper 402, European Central Bank

Campos, J., Ericsson N. and Hendry D. (2005) General-to-Specific Modeling: An Overview and Selected Bibliography, In J. Campos, D. F. Hendry, and N. R. Ericsson (Eds.), General-to-Specific Modeling, Volume 1. Cheltenham: Edward Elgar Publishing

Culver, S. and Papell, D. (1997) Is There a Unit Root in the Inflation Rate? Evidence from Sequential Break and Panel Data Models, Journal of Applied Econometrics, 12, 435-44

Demers, F and De Champlain, A. (2005) Forecasting Core Inflation in Canada: Should We Forecast the Aggregate or the Components?, Bank of Canada Working Paper 2005-44, December 2005

Den Reijer, A. and Vlaar, P. (2003) Forecasting inflation: An art as well as a science!, DNB Staff Reports, No. 107/2003

Espasa, A. and Albacete, R. (2004) Econometric Modelling for Short-Term Inflation Forecasting in the EMU, Statistics and Econometric Series, Working Paper 03-43

Fischer, S. (1977) Long-Term Contracts, Rational Expectations, and the Opti-

mal Money Supply Rule, Journal of Political Economy 85(1): 191–205

Friedman, M. The Role of Monetary Policy, American Economic Review, March 1968, 58, 1-17

Fuhrer, J. and Moore, G. (1995) Inflation Persistence, Quarterly Journal of Economics, 110(1): 127-60

Fuhrer, J., Little J., Kodrzycki Y. and Olivei G. (2009) Understanding Inflation and the Implications for Monetary Policy A Phillips Curve Retrospective, MIT Press

Gordon, R. (1977) The Theory of Domestic Inflation, American Economic Review 67(1): 128-34

Gordon, R. (1998) Foundations of the Goldilocks Economy: Supply Shocks and the Time-Varying NAIRU, Brookings Papers on Economic Activity 2: 297-333

Gordon, R. (2011) The History of the Phillips Curve: Consensus and Bifurcation, Economica, 78, 10–50

Gray, J. (1977) Wage Indexation: A Macroeconomic Approach, Journal of Monetary Economics 2(2): 221-35

Hendry, D. and Hubrich, K. (2010) Combining Disaggregate forecasts or combining disaggregate information to forecast an aggregate, Working Paper 1155, European Central Bank

Hubrich, K. (2005) Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?, International Journal of Forecasting, 21: 119-36

Lucas, R. (1972) Expectations and the neutrality of money, Journal of Economic Theory, 4, 103-24

Lucas, R. (1973) Some International Evidence on Output-Inflation Tradeoffs, American Economic Review 63(3): 326–34

Lucas, R. (1976) Econometric Policy Evaluation: A Critique, Carnegie- Rochester Conference Series on Public Policy, 1, 9–46

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

Muth, J. (1961) Rational Expectations and the Theory of Price Movements, Econometrica 29(3): 315–35

Phelps, E. (1967) Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time, Economica 34

Phelps, E. (1968) Money-Wage Dynamics and Labor-Market Equilibrium, Journal of Political Economy, July-August , 76, 678-711

Phillips, A. W. (1958) The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957, Economica 25(100): 283-99

Samuelson, P. and Solow, R. (1960). "Analytical Aspects of Anti-Inflation Policy," American Economic Review Papers and Proceedings 50(2): 177-94

Sargent, T. and Wallace N. (1975) "Rational" Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule, Journal of Political Economy 83(2): 241-54

Sargent, T. (1982) The ends of four big inflations. In R. E. Hall (ed.), Inflation. Chicago: University of Chicago Press, 41-97

Staiger, D., Stock, J. and Watson, M. (1997) The NAIRU, unemployment, and monetary policy Journal of Economic Perspectives, 11, 33–49

Stock, J. and Watson, M. (1996) Evidence on Structural Stability in Macroeconomic Time Series Relations, Journal of Business and Economic Statistics, 14: 11-30

Stock, J. and Watson, M. (1999) Forecasting inflation, Journal of Monetary Economics, 44(2): 293-335

Stock, J. and Watson, M. (2006) Forecasting with many predictors, Handbook of Economic Forecasting, Elliott, G., Granger, C.W.J., and Timmermann, A. (ed.), NorthHolland

Stock, J. and Watson, M. (2008) Phillips curve inflation forecasts. NBER Working Papers 14322, National Bureau of Economic Research, Inc

Taylor, J. (1980) Aggregate Dynamics and Staggered Contracts, Journal of Political Economy 88(1): 1-23