October 16, 2009

Forecasting Annual Inflation with Seasonal Monthly Data: Using Levels versus Logs of the Underlying Price Index

Helmut Lütkepohl[1]
Department of Economics, European University Institute
Via della Piazzola 43, I-50133 Firenze, ITALY
email: helmut.luetkepohl@eui.eu

Fang Xu
Max Weber Programme, European University Institute
Via delle Fontanelle 10, I-50014 San Domenico di Fiesole, ITALY
email: fang.xu@eui.eu

Abstract. This paper investigates whether using natural logarithms (logs) of price indices for forecasting inflation rates is preferable to employing the original series. Univariate forecasts for annual inflation rates for a number of European countries and the USA based on monthly seasonal consumer price indices are considered. Stochastic seasonality and deterministic seasonality models are used. In many cases the forecasts based on the original variables result in substantially smaller root mean squared errors than models based on logs. In turn, if forecasts based on logs are superior, the gains are typically small. This outcome sheds doubt on the common practice in the academic literature to forecast inflation rates based on differences of logs.

Key Words: Autoregressive moving average process, forecast mean squared error, log transformation, seasonally integrated process, seasonal dummy variables.
JEL classification: C22

[1]A preliminary version of this paper was presented at the conference on *Periodicity, Non-stationarity, and Forecasting of Economic and Financial Time Series*, Koldingfjord, Denmark, 21-23 August 2009. We thank the conference participants for valuable comments. We are also grateful to Denise Osborn, Andrea Silvestrini and Giacomo Sbrana for helpful comments. Financial support of the Fritz Thyssen Stiftung (Az.10.08.1.088) for the second author is also gratefully acknowledged.

1 Introduction

Forecasts of annual inflation rates are of great interest both for monetary policy and for planning purposes in various contexts in business and in economics more generally. A number of studies have explored possibilities for improving inflation forecasts. For example, Stock and Watson (1999) use the information in other economic variables to improve inflation forecasts. Espasa, Senra and Albacete (2002) and Hubrich (2005) use disaggregated data. Furthermore, Angelini, Henry and Mestre (2001), Camba-Méndez and Kapetaneous (2004) and Hofmann (2008) consider extracting useful information for forecasting inflation rates from large datasets. In all these studies inflation is measured by first or annual differences of logarithms (logs) of the underlying price index. In fact, using log price indices for inflation forecasting seems to be the leading approach in the academic literature, although this practice has not been properly justified.

In a recent study Lütkepohl and Xu (2009) investigated the role of the log transformation in forecasting economic time series. They found that using logs can bring about substantial gains in forecast accuracy if the log transformation leads to a more homogeneous variance of a variable. If this condition is not satisfied, it may be preferable to forecast the series directly. Therefore, in this study we investigate whether forecasting annual inflation rates on the basis of first or annual differences of logs is preferable to forecasting the underlying price indices directly. We focus explicitly on seasonally unadjusted price series because such series may have seasonal unit roots which can be captured by annual differences and seasonal adjustment brings about additional problems for modelling and forecasting.

Forecasting seasonal time series has been considered in a number of studies (e.g., Osborn, Heravi and Birchenhall (1999), Clements and Hendry (1997), Franses (1991), Paap, Franses and Hoek (1997), see also Hylleberg (1986) for a discussion of seasonality more generally). Osborn (2002) provides a review of the results. In these studies of seasonal data the question whether or not to use logs is either not explicitly explored at all or marginally touched upon. For example, Osborn et al. (1999) use some pre-screening of their series and decide on the log transformation based on that. Once the decision is made in favor of logs, all further modelling is based on the transformed series.

We compare univariate forecasts of inflation rates based on different consumer price indices for a range of countries. The mean squared error (MSE) or root mean squared error (RMSE) is used as the measure for forecast precision. We focus on autoregressive (AR) models based on seasonal differences, called *stochastic seasonality models*, and models with seasonal dummies for the first differences, called *deterministic seasonality models*. These two model classes were found to be more successful in forecasting seasonal time series than models with both first and seasonal differences and models without any differences at all (see in particular Osborn et al. (1999)). For comparison purposes we also use the so-called *airline model* for forecasting inflation rates. This model involves both first and seasonal differences and has a first order moving average (MA) and a seasonal MA part. Osborn (2002) suggests that this type of model may be worth considering for forecasting even if typical unit root tests do not support double differencing.

It is found that for inflation forecasting taking logs is by no means universally preferable to forecasting the price index directly. In fact, our results suggest that using the underlying price index directly should be the default in forecasting inflation. In many cases forecasting the price indices directly leads to substantially smaller forecast RMSEs while gains from using logs are usually slight. This result sheds doubt on the common practice of using prices in logs and constructing inflation rates as differences of log prices. It opens up another direction for forecast improvements.

Actually, the question whether it is useful to apply some kind of transformation and in particular the log transformation to a variable, prior to constructing a forecasting model, is an old one. Box and Jenkins (1976) discuss it already as part of their modelling strategy (see in particular their Section 9.3)[2] and Chatfield and Prothero (1973) apply logs to their sales series to make the variance more homogeneous when they construct a forecasting model. In fact, Tunnicliffe Wilson (1973) in his discussion of Chatfield and Prothero (1973) mentions that their use of logs may have led to poor forecasts and advocates the more general Box-Cox class of transformations (Box and Cox (1964)). We will not consider more general transformations in our forecast comparison because such transformations do not seem common in inflation forecasting. Clearly, in this context the log transformation has the advantage of giving rise to quantities with a direct interpretation. For instance, differences of log price indices are precise approximations of inflation rates. Moreover, in economic models log price indices are typically easy to interpret.

Our study is structured as follows. The forecasting models and forecasts are presented in the next section. In Section 3 a forecast comparison based on a range of consumer price index (CPI) series for different countries is presented. Section 4 concludes.

[2]The discussion of transformations has not changed much in the later edition of the book (Box, Jenkins and Reinsel (1994)).

2 Models and Methods

2.1 Variables of Interest

The price index of interest is denoted by y_t and its natural logarithm is signified as x_t, that is, $x_t = \log y_t$. It is assumed that y_t is observed s times per year. In the empirical study in Section 3 monthly series are considered, i.e., $s = 12$. For the discussion of models and methods we still prefer to entertain a more general setup which in principle also allows for other observation frequencies.

It is assumed that forecasts are desired for annual inflation rates based on the price index y_t,

$$\Delta_s y_t / y_{t-s},$$

rather than the approximate inflation rate $\Delta_s x_t$. Here Δ_s denotes the seasonal differencing operator defined such that $\Delta_s y_t = y_t - y_{t-s}$. The h-periods ahead forecasts of y_t and x_t at forecast origin t are denoted by $y_{t+h|t}$ and $x_{t+h|t}$, respectively. The forecast MSE or its square root (RMSE) are used as measures for forecast accuracy.

2.2 Forecasting Models

Different autoregressive (AR) models are fitted as forecasting models. For example, AR models may be considered for the seasonal differences of x_t and y_t, e.g.,

$$\Delta_s x_t = \nu + \alpha_1 \Delta_s x_{t-1} + \cdots + \alpha_p \Delta_s x_{t-p} + \varepsilon_t, \tag{1}$$

where ν is a constant term and ε_t is a white noise error term. Alternatively, a model with seasonal dummy variables may be fitted to the first differences,

$$\Delta x_t = \nu_1 \delta_{1t} + \cdots + \nu_s \delta_{st} + \alpha_1 \Delta x_{t-1} + \cdots + \alpha_p \Delta x_{t-p} + \varepsilon_t, \tag{2}$$

where $\Delta x_t = x_t - x_{t-1}$ and $\delta_{it} = 1$ if t is associated with the ith season (month) and $\delta_{it} = 0$ otherwise, that is, the δ_{it}'s are seasonal dummy variables. Model (1) is referred to as *stochastic seasonality model* and (2) is called *deterministic seasonality model*. We use full AR models with different lag orders p and subset AR models where some of the AR coefficients α_j, $j < p$, are zero. These models are fitted to x_t and y_t or, more precisely, the seasonal and first differences of x_t and y_t.

Notice that for monthly data Δx_t represents the monthly inflation rate from which an annual rate may be obtained as $s\Delta x_t$. We do not use this inflation measure because it puts the deterministic seasonality model at a severe disadvantage. Notice that this model is likely to produce seasonal monthly forecasts which need to be added to obtain a sensible annual forecast. Hence, we compute forecasts $x_{t+h|t} = \Delta x_{t+h|t} + \cdots + \Delta x_{t+1|t} + x_t$ for the log price index from the model for the first differences. These forecasts are then used to determine annual inflation forecasts as discussed in the following. Analogous comments apply for the forecasts based on models for Δy_t.

For comparison purposes we also consider forecasts based on the so-called *airline model* which is given by

$$\Delta\Delta_s x_t = \nu + (1 + \theta L)(1 + \theta_s L^s)\varepsilon_t, \tag{3}$$

where L denotes the backshift operator defined such that $L^i\varepsilon_t = \varepsilon_{t-i}$ and ν, θ and θ_s are parameters. In the classical Box-Jenkins terminology (Box and Jenkins (1976)) this model is a seasonal autoregressive integrated moving average (ARIMA)$(0,1,1)(0,1,1)_s$ model. Given that double differencing is used, a constant term may not be needed, that is, $\nu = 0$ may be assumed. We allow for a nonzero ν, that is, we estimate ν as an additional parameter, to avoid distortions of our forecasts due to a neglected nonzero mean term. The airline model is used for both x_t and y_t in the empirical section. It may also be summarized under stochastic seasonality models because it involves seasonal differences. We keep it separately because it involves MA terms. We estimate the airline model by a maximum likelihood (ML) procedure which conditions on zero initial values for the residuals and we restrict the seasonal MA coefficient to the interval $[-1, 0]$, that is, $\hat{\theta}_s \in [-1, 0]$ and the nonseasonal MA coefficient is restricted such that $\hat{\theta} \in [-1, 1]$. Thus, the seasonal MA part is set up to compensate for potential seasonal over-differencing whereas the nonseasonal MA term captures short-term correlation more generally. Restricting the seasonal MA parameter to the negative unit interval is in line with the procedure used in the SEATS seasonal adjustment programme (see Gómez and Maravall (1997)).

2.3 The Forecasts of Interest

The forecasts based on levels variables obtained from the airline model as well as the stochastic and deterministic seasonality models are denoted by $y^{air}_{t+h|t}$, $y^{ss}_{t+h|t}$ and $y^{ds}_{t+h|t}$, respectively. Forecasts for y_t are obtained from forecasts of x_t by reversing the log,

$$y^{nai}_{t+h|t} = \exp(x_{t+h|t}).$$

This notation was also used by Lütkepohl and Xu (2009) for a forecast for y_t obtained by applying the exponential function to reverse the log. Granger and Newbold (1976) pointed out that this forecast is in general not optimal and therefore they called it a naive forecast. If x_t is generated by a Gaussian linear process, e.g., by an ARIMA process with normally distributed white noise, the optimal forecast is

$$y^{opt}_{t+h|t} = \exp(x_{t+h|t} + \tfrac{1}{2}\sigma_x^2(h)),$$

where $\sigma_x^2(h)$ is the forecast error variance of an h-step forecast of x_t. Hence, the naive forecast is multiplied by an adjustment factor $\exp[\frac{1}{2}\sigma_x^2(h)]$ to obtain the optimal forecast. In our empirical analysis we also studied this predictor in preliminary investigations. It turned out, however, that it is at best marginally better than the naive forecast, that is, the resulting MSEs are often almost identical. Sometimes the optimal predictor even has a slightly larger MSE. This result is in line with an observation made by Lütkepohl and Xu (2009) in a univariate forecast comparison and by Bårdsen and Lütkepohl (2009) in a multivariate context. Three factors may contribute to this result. First, the forecast error variance is not known but has to be estimated. Thus, we just have an estimated adjustment factor for the naive forecast. Second, the normality assumption underlying the adjustment factor may be violated for some of our series. Third, the forecast error variance is typically very small compared to the level of the log price index. In that case the adjustment factor does not make much difference. As a consequence we do not consider this forecast further in the following.

In analogy with the previously introduced notation we denote the naive forecasts based on the airline model as well as the stochastic and deterministic seasonality models by $y^{air,nai}_{t+h|t}$, $y^{ss,nai}_{t+h|t}$ and $y^{ds,nai}_{t+h|t}$, respectively. Thus, in total we have six forecasts of the annual inflation rates based on a price index y_t which are compared subsequently:

$$\Delta_s y^{air}_{t+h|t}/y^{air}_{t+h-s|t}; \qquad \Delta_s y^{ss}_{t+h|t}/y^{ss}_{t+h-s|t}; \qquad \Delta_s y^{ds}_{t+h|t}/y^{ds}_{t+h-s|t};$$

$$\Delta_s y^{air,nai}_{t+h|t}/y^{air,nai}_{t+h-s|t}; \qquad \Delta_s y^{ss,nai}_{t+h|t}/y^{ss,nai}_{t+h-s|t}; \qquad \Delta_s y^{ds,nai}_{t+h|t}/y^{ds,nai}_{t+h-s|s}.$$

Note that in most of our comparisons we use $h \leq s$ so that the denominator is the observed value y_{t+h-s}.

2.4 Model Selection

Of course, it would be helpful to have criteria to decide a priori, for a given time series, which forecast is the best. One may conjecture that HEGY tests for seasonal unit roots (Hylleberg, Engle, Granger and Yoo (1990))[3] may be helpful in discriminating between stochastic and deterministic seasonality

[3]Extensions to monthly data are due to Franses (1990) and Beaulieu and Miron (1993).

models. If unit roots at all seasonal frequencies are found, fitting an AR model to the seasonal differences seems like a plausible strategy. Unfortunately, for the series considered in the next section HEGY tests typically do not reject seasonal unit roots at some frequencies but reject for others. A clear relation between the seasonal unit roots and the best forecasting model is not detected. Thus, it seems unwise to decide between the two model types on the basis of HEGY tests alone.

Also, of course, we have to choose between forecasts based on the original price indices and on the logs. We explore the potential of model selection criteria to help with the choice. In particular, we use the Akaike information criterion (Akaike (1973))

$$\text{AIC} = -2\log(\text{maximum likelihood}) + 2 \times \text{number of parameters}. \quad (4)$$

Note that the probability density function (pdf) of the sample $x_1, \dots, x_T$ denoted by $\phi_x(x_1, \dots, x_T)$ is related to the pdf of $y_1, \dots, y_T$ as

$$\begin{aligned} \phi_y(y_1, \dots, y_T) &= |\partial(x_1, \dots, x_T)/\partial(y_1, \dots, y_T)'|\, \phi_x(x_1, \dots, x_T) \\ &= \phi_x(x_1, \dots, x_T) \Big/ \prod_{t=1}^{T} y_t \ , \end{aligned}$$

where $|\cdot|$ signifies the determinant. To compare models for y_t and x_t via AIC, we use for both $\phi_x(x_1, \dots, x_T)$ and $\phi_y(y_1, \dots, y_T)$ Gaussian densities. The rational for this choice is that the Box-Cox transformation is often used to make the distribution more Gaussian. The transformation is defined as

$$y_t^{(\lambda)} = \begin{cases} (y_t^\lambda - 1)/\lambda & \text{for } \lambda > 0, \\ \log y_t & \text{for } \lambda = 0. \end{cases}$$

For $\lambda = 1$ this is just the original series shifted by -1, whereas for $\lambda \to 0$ the transformation approaches the log. Assuming that only $\lambda = 0$ and 1 are permissible parameter points, we may just compare a Gaussian likelihood for the original y_t's to that for the x_t's divided by $\prod_{t=1}^{T} y_t$. The comparison is done via AIC because models with different numbers of parameters may be used for y_t and x_t.

These considerations imply the following strategy for model selection. First AR models (1) and (2) are fitted to y_t and x_t. Then the four resulting models are compared with the AIC form in (4) taking into account the transformation of the distribution implied by the log transformation. In addition to the AIC we also use the Schwarz model selection criterion (Schwarz (1978)) which multiplies the number of parameters by $\log T$ instead of 2. It will be denoted by SC.

In the first step of this model selection strategy AR models are chosen in two alternative ways. More precisely, subset AR models specified by some subset selection procedure are considered and full AR models for which the lag length is selected by AIC or SC are employed. Subset AR models are specified by fitting an AR model of some maximum lag length and then successively deleting the coefficient with the smallest t-ratio, re-estimating the resulting model and deleting the next coefficient with the smallest t-ratio until all remaining t-ratios are larger than 1.96. Thus, in the final model all coefficients including the ones attached to deterministic terms are significant at a 5% level based on their t-ratios. Subset models are plausible in the present context because they tend to be more parsimonious than full AR models for seasonal monthly data. They typically choose small lags and lags around the seasonal period.

AIC and SC are also used for the airline model to choose between specifications in levels or logs. Given that the number of parameters is fixed, this procedure amounts to a comparison of the likelihood maxima. In fact, this decision rule appears to be the one used by the commonly applied seasonal adjustment procedures Tramo and Census X12 to choose between additive and multiplicative models (or, equivalently, levels and logs) for a given time series (see EViews (2000)).[4] The algorithms used for estimating the parameters of the airline model in these two procedures differ from ours, however. They both use exact ML procedures based on different algorithms. Moreover, Tramo and Census X12 do not seem to include an intercept term in the airline model. Hence, their estimates of the MA coefficients may differ from ours.

3 Forecast Comparison for Inflation Rates

3.1 The Data

We use two sets of unadjusted monthly CPI data to compare forecasts and investigate procedures for choosing a good forecasting model. The first one consists of total CPI series for 24 countries for the period 1996M1-2007M12. The countries are member states of the European Union (EU) and most but not all of them are members of the euro area, i.e., the European Monetary

[4]Although the Tramo-Seats manual discusses a possibility to choose between levels and logs on the basis of a range-mean regression (Gómez and Maravall (1997, p. 15)), A. Maravall, in a personal communication, confirmed that the choice is actually made by a likelihood comparison in more recent versions of the program.

Union (EMU). In addition, we also include data for the USA.[5] The sample period is chosen to account for the special situation in Europe during the run-up to the EMU. The euro was introduced in January 1999 and most countries had to adjust their inflation rates to satisfy the Maastricht criteria. Therefore inflation rates in many European countries have changed during the 1990s. The process may not have been completed in 1996. We still start our sample in this year to ensure a reasonably long sample period for model specification and estimation in our setup where the last years are reserved for the forecast comparison. We have checked the impact of the first three years on our results and report on the outcome later. We also have data for the year 2008 but do not report results based on them. The year 2008 is a bit special in that the inflation rates in most countries were rather high at the beginning and then dropped sharply. Clearly such an unusual period may cause difficulties for forecasting methods. Some poor forecasts may dominate the MSE if the unusual period is included. In particular, when inflation is high the inflation rates computed as differences of logs will be poor approximations to the actual inflation rates. Thus, including data from 2008 might have biased our results against logs.

Plots of the series are shown in Figure 1. The last four series from Belgium, Italy, Luxembourg and Spain have an obvious change in their seasonal pattern during our sample period. The problem becomes even more apparent in Figure 2, where the first differences of the four series are plotted. Although a change in seasonality can be captured to some extent by a stochastic seasonality model, it is difficult to argue that these series can be classified as difference stationary. We still keep them in our sample to study the impact on our forecast comparison. All the other series have a more or less regular seasonal pattern over our sample period. At least a visual inspection does not give rise to concerns about structural breaks.

In fact, there have been adjustments to the construction of the indices during our sample period. In particular, sales prices have been included in the HICPs published by Eurostat.[6] These alterations explain the apparent structural changes in the series for Belgium, Italy, Luxembourg and Spain. They may also have affected the price indices of other countries. We ignore this problem because it is not obvious from a visual inspection of the series. In fact, Bataa, Osborn, Sensier and van Dijk (2008) found evidence for structural change in the price indices of other series as well. Again we ignore such changes because they are not apparent in the graphs of the series. Obviously,

[5]The data for the EU countries are the HICP series from Eurostat taken from the database of the European Central Bank. The US series is total CPI from the database of the Federal Reserve Bank of St. Louis.

[6]We thank Denise Osborn for drawing our attention to this adjustment.

our focus is not on possible structural changes in the series although they may admittedly be important for the forecast performance of our models. We have no reason to believe, however, that they affect models for levels more or less than models for logs. Hence, we ignore possible problems due to structural changes which are not apparent in a visual inspection of the series.

For some of the series the seasonal fluctuations increase with the level of the series (see, e.g., the series from Malta or the USA). Thus, the volatility increases with the level. This behavior can potentially be alleviated by a log transformation. On the other hand, for some of the other series applying the log transformation does not make much difference for their general appearance (e.g., for Austria, Cyprus, Denmark, Sweden or the UK).

In Table 1 the results of standard augmented Dickey-Fuller (ADF) and HEGY tests for seasonal unit roots are reported. Assuming that all the series have at most one unit root and applying ADF tests to the original series and their logs, it turns out that only four rejections are obtained at the 5% level of significance. More precisely, unit roots are rejected for the original series for Cyprus and Malta and for the logs of the Hungarian and Polish series. Thus, for the vast majority of the series there is some evidence that taking at least first differences is justified for the levels as well as the logs. Of course, in our forecast comparison we treat all series in the same way as far as differencing is concerned. It may be useful, however, to keep in mind the results of the unit root tests.

When ADF tests are applied to seasonal differences, unit roots are rejected in nine cases. For Cyprus, the Czech Republic and Greece unit roots are rejected for both the original series and the logs, whereas for Malta the ADF test rejects for the levels and for Poland and Belgium for the logs. Overall there is some evidence for unit roots in the seasonal differences of most series. Given the results for the logs, this means, of course, that the ADF tests indicate a unit root in the annual inflation rates of most countries.

On the other hand, when HEGY tests are applied to the first differences, for all countries with a single exception, at least some seasonal unit roots are rejected. The exceptional case is Malta. For all the other countries taking both first and seasonal differences is, hence, not supported by a HEGY test. Although this sheds doubt on the suitability of the airline model, we nevertheless use it for comparison purposes.

To investigate the robustness of our results we also consider some other price indices. In particular, our second data set consists of 16 core CPI series. We have comparable data for all countries from the previously described set except for the Czech Republic, Hungary and Slovenia and we have eliminated the series from Belgium, Italy, Luxembourg, Portugal and Spain because of obvious structural breaks. The sample period is again 1996M1-2007M12 and the data sources are the same as for the total CPI series. Our core CPI series for European countries are the HICPs of all items excluding energy and food and the corresponding CPI is used for the USA.

The unit root properties are similar to those of the total CPI series. More precisely, a zero frequency unit root is diagnosed in all series and their logs except for logs of the series from Greece. Furthermore a zero frequency unit root is not rejected for all but three seasonally differenced series, the exceptions being levels and logs for Germany and logs for Greece. Seasonal unit roots are found at best for some of the seasonal frequencies in the first differences of levels and logs of the series from all countries. Again these results suggest that applying first and seasonal differences may lead to over-differencing. The core inflation series are primarily used for additional robustness checks. Therefore we do not enter into a more detailed discussion of their properties at this stage.

Core inflation series are of interest to policy makers because they eliminate some more volatile inflation components. For our purposes it may also be of interest to investigate whether our results hinge on considering rather smooth and regular price indices, such as total and core CPI. The related inflation rates were relatively low during the forecast periods. Therefore, we also consider a set of more volatile price indices for energy. More precisely, we use the HICP series for energy from all European countries from the total HICP panel except for Czech Republic, Hungary, Malta and Slovenia, which leaves us with 19 series. The data are originally from Eurostat, as given in the database of the European Central Bank. The seasonal component in some of these series is less apparent than in those of the other datasets. It is to some extent dominated by general volatility. Note that some related energy inflation rates were substantially in excess of 10% during parts of our sample period. Thus, they are at times much larger than total inflation.

In our forecast comparison we first focus on the total CPI series and then study the other inflation series.

3.2 Baseline Forecast Comparison for Total CPI Series

We set aside the observations for the last two years of the sample for forecasting and fit the airline model as well as the stochastic and deterministic seasonality models to all total CPI series and their logs. For the 1-step forecasts the smallest specification and estimation period is 1996M1-2005M12, whereas for the 12-step forecasts it is 1996M1-2005M1. The following conclusions emerge from Table 2. The AR models are subset models chosen with the lag elimination procedure outlined in Section 2.4 with a maximum AR order of 14. The six forecasts and associated forecast errors are computed on the basis of these models. Then the sample is extended by one observation and the model specification and estimation steps are repeated and so on. Thereby we get a set of forecast errors on which the RMSEs are based.

The best forecasting models with the smallest RMSEs within each model class for 1-step and 12-step forecasts are given in Table 2. Significant differences of the related MSEs at a 5% level between models in levels and logs according to the Harvey, Leybourne and Newbold (1997) version of the Diebold-Mariano test (Diebold and Mariano (1995)) are indicated by an asterisk. Also the overall best models for each of the two forecast horizons are presented in the table. The table is based on the RMSEs obtained over 24 forecasts for the last two years of our sample.

1. Considering only the 20 series without obvious structural breaks (the first 20 in the table), forecasts based on levels are optimal in about half of the cases. In fact, considering the different model types separately, levels are preferable in a clear majority of the cases for the airline and stochastic seasonality models. These results hold for both forecast horizons. In a substantial number of cases the levels forecasts are in fact significantly better than the forecasts based on logs whereas the converse is true only in three cases. As mentioned earlier, the seasonal pattern for many of the countries does not depend much on the level of the series (see also Figure 1). Hence, taking logs does not make them more regular which may explain the good performance of levels forecasts.

2. Again for the 20 series without apparent structural breaks, the deterministic seasonality and stochastic seasonality (including airline) models are optimal for roughly similar numbers of countries. For 1-step forecasts seven out of 20 times the deterministic seasonality model is overall best while the corresponding number for 12-step forecasts is ten.

3. The ranking of the models with respect to their forecast performances depends to some extent on the forecast horizon. For example, for Denmark a deterministic seasonality model delivers the best 12-step forecasts while an airline model (that is, a stochastic seasonality model) is best for 1-step forecasting. The reverse outcome is observed for Finland.

4. The airline model performs overall better for the original data than for the logs, both for one month and one year ahead forecasts. It is not very successful compared to the AR models for 12-step forecasts,

however. For this forecast horizon it comes out best in only five cases. Perhaps this result reflects the message given by the HEGY tests which did not support double differencing.

5. For the series with a break in the seasonal pattern (the last four in Table 2) the overall results are not much different from those for more regular series.

In summary, these observations mean, of course, that the common practice of basing annual inflation forecasts on first and/or seasonal differences of logs is questionable. There is some potential for forecast improvements at least in some countries by using the original levels series. Of course, the actual gains in forecast precision cannot be seen from Table 2 even though in some cases the differences between forecasts based on levels and logs are significant. To show that levels forecasts can lead to substantial improvements, we present relative RMSEs in Figure 3 for 12-step forecasts. The longer forecast horizon is more relevant from a policy point of view and, hence, we present detailed results for that. All RMSEs are divided by the RMSE of the corresponding airline model based on the original variables, that is, the latter model is the benchmark model in Figure 3. A black bar in each panel signifies the smallest RMSE. The results in the figure show that the differences between the models can indeed be substantial (see, e.g., Cyprus, Hungary, Ireland, Malta or the UK). Similar results are also found for 1-step forecasts. Thus, using the original levels of some of the series leads potentially to much more precise forecasts than models based on logs.

These conclusions raise the question of how to find the best model for a given series for out-of-sample forecasting. In particular, the question in which cases levels should be used to improve inflation forecasts is of considerable interest. To investigate this question we present the recommendations of several model choice criteria and procedures in Table 3. We report the recommendations of the AIC and SC selection procedures and compare our airline models for levels and logs based on AIC (or, equivalently, the corresponding likelihood maxima). Note that our airline model used in the comparison with the subset AR models is estimated on the same sample as the subset AR models to ensure a fair comparison via the model selection criteria. In other words, the first 14 values which are used as presample values for model selection for the subset AR models are also not considered for estimating the airline model. However, the column with the heading "Airline" in Table 3 contains the recommendations based on the airline model when estimated over the full sample to have a comparable procedure to those implemented in the TRAMO and Census X12 programs. Note that in TRAMO the airline model is used as the standard model for choosing between levels and logs while in the Census X12 program an ARIMA model can be selected which we have specified to be an airline model. The recommendations of these procedures are also shown in Table 3 and they are also estimated using the whole sample. In Table 3 we present results for varying sample periods to check the robustness of the results. The following conclusions emerge:

1. Unfortunately, the recommendations in the table are not reliable indicators of the best forecasting model. For example, if only data before our first forecast period are considered (1996M1-2004M12), the best models according to AIC agree only in 6 cases with the overall best 12-step forecasting models in Table 2. Also SC is not a reliable indicator for choosing the best forecasting model.

2. None of the procedures is fully reliable in choosing between levels and logs for forecasting purposes. Focussing again on the 12-step forecasts in Table 2, it turns out that the recommendation from the Census X12 procedure agrees in most cases with the best forecasting model. Its recommendation conforms in 18 out of 24 cases with the best choice for forecasting. Even that leaves room for improvement, of course. It may also be worth emphasizing that forecasting is not the main objective of the TRAMO and Census X12 programs.

3. The different criteria typically do not agree on which form of a particular CPI should be used (levels or logs). In fact, only in very few cases is the same recommendation given by all criteria. Even then it is not clear that this recommendation corresponds to the best forecasts. For example, for Finland the log is unanimously recommended by all procedures and across all sample periods in Table 3. In contrast, the best 12-step forecasts in Table 2 are obtained from a model for the levels. Note, however, that for Finland the RMSE differences for levels and logs are not significant.

4. The recommendations for a single series frequently change with the sample period. In other words, by adding a couple more years to the sample, a given procedure may reverse its decision regarding levels or logs of a variable. For example, the Census X12 procedure changes its recommendation in six cases when a sample period from 1996M1-2007M12 is used instead of 1996M1-2004M12, that is, when only three years of additional data are added. Thus, the decisions are rather fragile.

5. Although TRAMO and Census X12 are both based on airline models they do not always agree among each other and with the choice based

on our airline model. The reason is that there are slight differences in the three procedures, as described in Section 2.4. Apparently, these differences are enough to end up with different recommendations regarding levels or logs. Even for those cases where the three procedures agree, this is no insurance for a good choice for forecasting purposes.

The overall conclusion from the results in Table 3 is that the procedures in current use can at best give indications but cannot be used as reliable indicators for a decision on using levels or logs of a price index for forecasting purposes.

One may conjecture that deciding between levels and logs on the basis of past forecast performance may be a good strategy. Therefore we present forecasting results for a sample ending in 2005M12 in Table 4. These are obtained analogously to the results in Table 2. In other words, the forecasts are based on the data from the last two years of the sample. Comparing these results to those in Table 2 shows that there are many changes in the optimal models, overall and also within each model class. For example, the best airline model for 1-step forecasting in Table 2 is based on levels in 16 cases while the corresponding number in Table 4 is only 5. Even if we consider the overall best models for 12-step forecasting, they change for the vast majority of the countries in our sample when the sample ends in 2005M12 rather than 2007M12. Thus, if one chooses the forecasting model on the basis of data up to 2005, the model is not optimal for predicting inflation for the following years for most countries. Even the choice between levels and logs based on past forecasting performance is not optimal in many cases. This result also implies that in-sample methods cannot be expected to provide a reliable decision rule between levels and logs.

The overall conclusion is that computing inflation forecasts from the original price levels leads to substantially better forecasts for some countries and forecast horizons than using the log price indices. This result was found within the different model classes and overall across all model classes. Because it depends on the forecast horizon, the sample and forecast periods whether levels or logs are superior, in-sample methods cannot be expected to indicate whether levels or logs are preferable for forecasting. Hence, the sizable gains from using levels obtained in some cases suggest that using levels of price indices for forecasting inflation rates should be the default rather than using logs.

3.3 Robustness Analysis

We have checked the robustness of these results by varying the sample period, the model types, the model selection strategy, the forecast horizon and the types of price indices. In view of the start of the EMU in January 1999 we have eliminated the first three years of our sample and have repeated the comparison for the core CPI series. Results are presented in Table 5. Our main conclusion that using levels instead of logs is preferable for many countries is reinforced by these results. Again, forecasts based on logs are rarely significantly better than the corresponding forecasts based on levels.

Although parsimonious models are often preferable for forecasting purposes we have also fitted full AR models choosing only the lag length but not eliminating intermediate lags. The results are shown in Tables 6 and 7 for lag order selection by AIC and SC, respectively. Both tables convey the same picture, namely that using levels rather than logs is often beneficial for forecasting purposes. Using logs is rarely significantly better. In fact, for our series the full AR models often select lag orders around the seasonal lags and, hence, are not very parsimonious. They still produce better forecasts than the subset models in many cases. This can be seen in Figure 4, where 12-step forecasts based on deterministic seasonality models for the subset AR and full AR models based on AIC and SC are compared. It can be seen that only for five out of 24 countries the subset AR models produce better forecasts than the full AR models.

We have also computed biannual forecasts, that is, 24-step forecasts, to study the impact of a longer forecast horizon on our results although one could argue that univariate time series models may not be the best tools for longer term forecasting. Again the results are similar to those for the shorter horizons and are therefore not presented in detailed tables or graphs. The bottom line is again that in many cases level forecasts dominate forecasts based on logs although there is some variation relative to the shorter term forecasts. For example, there are cases where 12-step forecasts based on levels are better, while 24-step forecasts based on logs produce a smaller RMSE for the same variable. The reverse outcome is also observed. Thus, the 24-step forecasts do not change our general conclusions.

As a further robustness check we now turn to the other inflation series. For the core CPI series based on a sample period 1996M1-2007M12 similar conclusions emerge as for the total CPI series (detailed results are not shown). In particular, for 12-step forecasts levels are better than logs in eight out of 16 cases. With three exceptions, only level forecasts are significantly preferable when level and log forecasts are compared for specific models. For 1-step forecasts the results are even more in favor of levels forecasts. For 9 of the

16 countries levels produce better forecasts. Thus, for a majority of the countries there is some potential to improve forecasts by using levels rather than logs. Unfortunately, for a single country for which levels are better for predicting total CPI inflation, logs may be better for core inflation and vice versa. This result suggests that knowing more about the economic structure of a country and its specific seasonal pattern of the CPI may not be sufficient to decide on the best forecasting model. Similar results emerged when using full AR models or when using a reduced sample size by focussing on the period 1999M1-2007M12.

For the more volatile energy price indices and corresponding inflation rates the situation is not much different. Using again a sample from 1996M1-2007M12, levels 1-step and 12-step forecasts are optimal for 14 and 9 out of 19 countries, respectively. Hence, we can conclude that for a large range of price indices using levels improves forecasts at relevant horizons relative to using logs.

4 Conclusions

Given the importance of inflation forecasting, various proposals have been made in the recent literature to improve inflation forecasts. For example, the possibility of using disaggregated price series or incorporating information from other related series has been explored. In this study we have focussed on the question whether levels or logs of the underlying price index should be used for forecasting purposes. In the academic literature on inflation forecasting, using logs seems to be preferred although there does not seem to be a systematic investigation of the transformation issue. We have based forecasts on a number of univariate time series models which are plausible for seasonal monthly price series. More precisely, we have used the so-called airline model which is based on first and seasonal differences and we have also used AR models for first differences and for seasonal differences. That is, we have employed deterministic and stochastic seasonality models.

Our forecast comparison is based on different monthly seasonal CPI series from a large number of European countries and the USA for the period 1996M1-2007M12. The results clearly show that the common practice of using log differences to approximate the inflation rate is not necessarily optimal. In fact, for a number of our series sizable and statistically significant forecast improvements are obtained by modelling the original CPI series rather than its log. On the other hand, if forecasts based on logs are superior, the gains are usually small. Hence, our results suggest that using levels should be the default.

Unfortunately, we have not found a reliable in-sample method to indicate for a given series whether levels or logs will lead to better out-of-sample forecasts. We have explored the ability of standard methods to help in this decision as they are implemented in the commonly used seasonal adjustment procedures based on Tramo and Census X12. While these methods can help they are certainly not very reliable indicators for deciding between levels and logs. In fact, our results do not give hope that such methods can be reliable tools for this purpose. The reason is that whether levels or logs are optimal for forecasting depends very much on the sample, forecast periods and forecast horizons. If additional data are added to the series of a particular country for which using logs is optimal for the shorter period, levels may be preferable for the longer period. Moreover, if levels produce better 1-step forecasts, logs may be better for 12-step forecasts and vice versa. Thus, it is not likely that in-sample procedures for discriminating between levels and logs can reliably choose the best form of the data.

Given the potential forecast improvements from using levels rather than logs, it may be worth exploring whether alternative methods for improving inflation forecasts such as using disaggregated price indices or information from other variables can benefit from allowing for the possibility of modelling levels rather than logs. Our study also suggests a number of other directions for further research. We have considered a range of different CPIs for a large set of European countries and the USA. In future research it may be of interest to explore the situation for countries from other regions of the world where inflation has had quite different characteristics. Moreover, using other forecasting models including multivariate ones may give further useful insights regarding the question whether logs should be used or not for particular series. Also, allowing for the more general Box-Cox transformation rather than using logs may be an interesting strategy for future research.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *in* B. Petrov and F. Csáki (eds), *2nd International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267–281.

Angelini, E., Henry, J. and Mestre, R. (2001). Diffusion index-based inflation forecasts for the euro area, *Working Paper series No 61*, European Central Bank.

Bårdsen, G. and Lütkepohl, H. (2009). Forecasting levels of log variables in vector autoregressions, *Working Paper EUI ECO 2009/24*, European University Institute.

Bataa, E., Osborn, D. R., Sensier, M. and van Dijk, D. (2008). Identifying changes in mean, seasonality, persistence and volatility for G7 and euro area inflation, *Discussion Paper 109*, University of Manchester.

Beaulieu, J. J. and Miron, J. A. (1993). Seasonal unit roots in aggregate US data, *Journal of Econometrics* **55**: 305–328.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society B* **26**: 211–243.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control, Third Edition*, Prentice-Hall, Upper Saddle River, NJ.

Camba-Méndez, G. and Kapetaneous, G. (2004). Forecasting euro area inflation using dynamic factor measures of underlying inflation, *Working Paper series No 402*, European Central Bank.

Chatfield, C. and Prothero, D. L. (1973). Box-Jenkins seasonal forecasting: Problems in a case-study (with discussion), *Journal of the Royal Statistical Society A* **136**: 295–336.

Clements, M. P. and Hendry, D. F. (1997). Empirical study of seasonal unit roots in forecasting, *International Journal of Forecasting* **13**: 341–355.

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

Espasa, A., Senra, E. and Albacete, R. (2002). Forecasting inflation in the European Monetary Union: A disaggregated approach by country and by sectors, *European Journal of Finance* **8**: 402–421.

EViews (2000). *EVievs 4.0 User's Guide*, Quantitative Micro Software, Irvine, CA.

Franses, P. H. (1990). Testing for seasonal unit roots in monthly data, *Econometric Institute Report 9032A*, Erasmus University Rotterdam.

Franses, P. H. (1991). Seasonality, non-stationarity and the forecasting of monthly time series, *International Journal of Forecasting* **7**: 199–208.

Gómez, V. and Maravall, A. (1997). Programs Tramo and Seats: Instructions for the user, *manual*, Madrid.

Granger, C. W. J. and Newbold, P. (1976). Forecasting transformed series, *Journal of the Royal Statistical Society B* **38**: 189–203.

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

Hofmann, B. (2008). Do monetary indicators lead euro area inflation?, *Working Paper series No 867*, European Central Bank.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?, *International Journal of Forecasting* **21**: 119–136.

Hylleberg, S. (1986). *Seasonally in Regression*, Academic Press, Orlando.

Hylleberg, S., Engle, R. F., Granger, C. W. J. and Yoo, B. S. (1990). Seasonal integration and cointegration, *Journal of Econometrics* **44**: 215–238.

Lütkepohl, H. and Krätzig, M. (eds) (2004). *Applied Time Series Econometrics*, Cambridge University Press, Cambridge.

Lütkepohl, H. and Xu, F. (2009). The role of the log transformation in forecasting economic variables, *Working Paper MWP2009/06*, European University Institute.

Osborn, D. R. (2002). Unit-root versus deterministic representation of seasonality for forecasting, *in* M. P. Clements and D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 409–431.

Osborn, D. R., Heravi, S. and Birchenhall, C. R. (1999). Seasonal unit roots and forecasts of two-digit European industrial production, *International Journal of Forecasting* **15**: 27–47.

Paap, R., Franses, P. H. and Hoek, H. (1997). Mean shifts, unit roots and forecasting seasonal time series, *International Journal of Forecasting* **13**: 357–368.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**: 461–464.

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

Tunnicliffe Wilson, G. (1973). Discussion of paper by Dr Chatfield and Dr Prothero, *Journal of the Royal Statistical Society A* **136**: 315–319.

Table 1: Unit Root Properties of Total CPI Series for Sample Period 1996M1-2007M12

	ADF with trend and seas. dummies		ADF with constant		roots not rejected by HEGY	
Country	level	log	Δ_{12} level	Δ_{12} log	Δ level	Δ log
Austria	-0.57 (0)	-1.04 (0)	-0.74 (14)	-1.09 (14)	1,2,3/4 (2)	1,2,3/4 (0)
Cyprus	-3.44* (9)	-3.37 (9)	-2.98* (12)	-3.05* (12)	1,2,3/4 (1)	1,2,3/4 (1)
Czech Rep.	-2.66 (6)	-2.75 (6)	-3.36* (12)	-4.02* (12)	1,2,3/4 (1)	1,2,3/4 (2)
Denmark	-2.15 (0)	-1.88 (0)	-2.84 (12)	-2.74 (12)	all but 9/10(12)	1,2,3/4 (0)
Finland	-2.98 (13)	-2.84 (13)	-2.22 (12)	-2.21 (12)	1,2,3/4 (0)	1,2,3/4 (0)
France	-1.30 (2)	-1.57 (2)	-1.36 (12)	-1.62 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Germany	0.16 (1)	-0.30 (1)	-1.04 (12)	-1.40 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Greece	-1.40 (12)	-3.23 (12)	-3.22* (0)	-3.58* (0)	1,3/4 (0)	1,3/4 (0)
Hungary	-2.20 (1)	-3.85* (6)	-2.45 (13)	-2.72 (13)	1,2,3/4 (0)	1,2,3/4,7/8 (0)
Ireland	-2.60 (0)	-1.68 (1)	-2.07 (12)	-1.89 (12)	1,2,3/4 (6)	1,2,3/4 (6)
Lithuania	-0.62 (6)	-1.11 (6)	-0.48 (14)	-0.88 (14)	1,2,3/4 (2)	1,2,3/4 (2)
Malta	-4.45* (6)	-2.54 (9)	-2.96*(12)	-2.54 (12)	all (12)	1,2,3/4 (0)
Netherlands	-1.51 (12)	-2.59 (12)	-1.79 (13)	-1.61 (13)	1,2,3/4 (0)	1,2,3/4 (0)
Poland	-2.35 (1)	-3.46* (1)	-2.46 (13)	-3.01* (13)	1,2,3/4 (0)	1,2,3/4,7/8 (0)
Portugal	-2.44 (6)	-2.04 (6)	-1.84 (12)	-1.79 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Slovakia	-1.13 (0)	-0.16 (0)	-1.70 (12)	-1.12 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Slovenia	-3.01 (14)	-1.81 (13)	-1.63 (12)	-1.55 (12)	1,2,3/4,7/8,11/12 (1)	1,2,3/4 (0)
Sweden	-1.95 (0)	-2.02 (0)	-1.61 (12)	-1.75 (12)	1,2,3/4 (0)	1,2,3/4 (0)
UK	0.45 (0)	-0.14 (0)	-0.58 (12)	-0.94 (12)	1,2,3/4 (0)	1,2,3/4 (0)
USA	-1.05 (12)	-1.85 (12)	-1.08 (14)	-2.01 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Belgium	-1.18 (6)	-1.64 (6)	-2.84 (11)	-3.22* (11)	1,2,3/4 (0)	1,2,3/4 (0)
Italy	-3.00 (12)	-3.00 (12)	-1.68 (12)	-2.02 (12)	all but 5/6 (0)	all but 5/6 (0)
Luxembourg	-1.34 (7)	-1.78 (7)	-1.86 (12)	-2.25 (12)	1,2,3/4 (0)	1,2,3/4 (0)
Spain	-1.82 (10)	-2.48 (10)	-1.28 (12)	-1.86 (12)	1,2,3/4 (7)	1,2,3/4 (3)

Note: Lag selection by AIC with maximum order 14, lag order given in parentheses. 5% critical values for ADF test: -3.41 (with trend), -2.86 (with constant). HEGY test with seasonal dummies, results based on 5% significance level. Computations were performe with JMulTi (Lütkepohl and Krätzig (2004)).

Table 2: Forecasting Results for Total CPI Series for Sample Period 1996M1-2007M12

	best 1-step forecast				best 12-step forecast			
Country	airline	ss	ds	overall	airline	ss	ds	overall
Austria	level	level	log	ds+log	log	log	log	airline+log
Cyprus	level	level	level	ds+level	level	level*	level	ds+level
Czech Rep.	level*	level	level	ds+level	level*	level	level	ds+level
Denmark	level	log	log	airline+level	log	level	log	ds+log
Finland	level	level	level	ds+level	level	level	log	airline+level
France	level	level	log*	airline+level	log	level*	log	ds+log
Germany	level	level	log	airline+level	log	log	log	airline+log
Greece	level	log	level	ss+log	level	log	log	ss+log
Hungary	level	level	level	airline+level	level	level	level	ds+level
Ireland	level	level*	level	ds+level	level	level*	level	ss+level
Lithuania	level	level*	level	airline+level	level*	level	level*	ds+level
Malta	log	level	level	ds+level	log	level	level	ss+level
Netherlands	log	log	level	ss+log	log	level	level*	ds+level
Poland	level	level	log	ss+level	level*	log	log	ds+log
Portugal	log	level	log	airline+log	log	level*	log	ds+log
Slovakia	level	level	level	ss+level	level*	log*	level	ss+log
Slovenia	level	level	log	ss+level	level	level	level	airline+level
Sweden	level	level	log	ds+log	level	level	log	ds+log
UK	log	level	log	airline+log	log*	level*	log	ss+level
USA	log	log	log	airline+log	log	log	log*	airline+log
Belgium	level	log	log	ds+log	level	log	log	ss+log
Italy	log	log	level	airline+log	level	log	log	ss+log
Luxembourg	log	log	level	ds+level	log	log	log	ds+log
Spain	log	log	log	ss+log	level	log	level	ds+level

Note: ss - stochastic seasonality subset AR model, ds - deterministic seasonality subset AR model. The underlying RMSEs are based on 24 forecasts. An asterisk (*) indicates a significant improvement according to a 5% level modified Diebold-Mariano test as proposed by Harvey et al. (1997).

Table 3: Best Performing Models In-Sample for Total CPI

Country	Period:1996M1-2004M12					Period:1996M1-2007M12				
	AIC	SC	Airline	Tramo	X12	AIC	SC	Airline	Tramo	X12
Austria	ds+level	airline+level	level	level	log	ds+log	airline+log	log	log	log
Cyprus	ds+log	ds+level	level	level	level	ds+level	ds+level	log	level	level
Czech Rep.	ds+log	ds+log	log	level	level	ds+log	ds+log	level	level	level
Denmark	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log
Finland	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log
France	ds+level	ds+level	level	level	log	ds+log	airline+log	level	log	log
Germany	ds+log	ds+log	level	level	log	ds+log	ds+log	level	level	log
Greece	ds+log	airline+log	log	log	log	ds+level	airline+log	log	log	level
Hungary	ds+level	ds+log	level	level	level	ds+log	ds+log	level	level	level
Ireland	ds+log	ds+log	log	log	log	ds+log	ds+level	level	log	log
Lithuania	ds+level	ds+level	log	level	level	ds+level	ds+log	level	level	level
Malta	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log
Netherlands	ds+level	ds+level	level	level	level	ds+level	ss+level	level	level	level
Poland	ds+level	ds+level	level	level	level	ds+level	ds+level	level	level	level
Portugal	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log
Slovakia	ds+level	ds+level	level	log	log	ds+level	ds+level	level	level	level
Slovenia	ds+log	ds+level	level	log	level	ds+log	ds+log	log	log	log
Sweden	ds+log	ds+log	level	log	log	ds+log	ds+log	level	log	log
UK	ds+level	ds+level	level	level	level	ds+log	airline+level	level	level	log
USA	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log
Belgium	ds+level	ds+level	level	level	log	ds+log	ds+level	log	level	level
Italy	ds+log	ds+log	level	log	log	ds+log	ds+log	level	log	log
Luxembourg	ds+level	ds+level	log	level	level	ds+level	ds+level	level	log	log
Spain	ds+log	ds+log	log	log	log	ds+log	ds+log	log	log	log

Note: ss - stochastic seasonality subset AR model, ds - deterministic seasonality subset AR model.

Table 4: Forecasting Models for Total CPI Series for Sample Period 1996M1-2005M12

Country	best 1-step forecast				best 12-step forecast			
	airline	ss	ds	overall	airline	ss	ds	overall
Austria	log	level	level	airline+log	log	level	log	airline+log
Cyprus	log	level	level	airline+log	log	level	level*	ds+level
Czech Rep.	log	level	log	ss+level	level*	level	log	ds+log
Denmark	log	level	log	ds+log	level	log	level	ds+level
Finland	log	log	level	airline+log	level	log	level	ds+level
France	level	log	log	ds+log	level	level	level	airline+level
Germany	log	level	log	airline+log	level	log*	log*	ss+log
Greece	level	log	level	airline+level	level*	log	log	ds+log
Hungary	level	level	level	airline+level	level	level	level	airline+level
Ireland	level	level	log	airline+level	level*	level*	level*	ds+level
Lithuania	log	log	level	ds+level	log	log	level	ds+level
Malta	log	level	log	airline+log	level*	level	level	ss+level
Netherlands	log	level	level	airline+log	log	level	level	ds+level
Poland	log	level	level	ds+level	level*	level*	level	ss+level
Portugal	log	level	log	ss+level	level	level	level	ds+level
Slovakia	log	level	level	ds+level	level	level	level	ds+level
Slovenia	log	level	level	airline+log	log*	level	level*	ss+level
Sweden	log	log	level	airline+log	level	level	level*	ds+level
UK	level	level	level*	ds+level	level*	log	log	ss+log
USA	log	log	log	ss+log	level	log	log	airline+level
Belgium	log	log	log	ds+log	log	log	log	ss+log
Italy	log	log	log	ss+log	level	level	level	ds+level
Luxembourg	log	level	log	airline+log	level	level	level	airline+level
Spain	log	log	log	ds+log	level	level	log	airline+level

Note: ss - stochastic seasonality subset AR model, ds - deterministic seasonality subset AR model. The underlying RMSEs are based on 24 forecasts. An asterisk (*) indicates a significant improvement according to a 5% level modified Diebold-Mariano test as proposed by Harvey et al. (1997).

Table 5: Forecasting Models for Total CPI Series for Sample Period 1999M1-2007M12

	best 1-step forecast				best 12-step forecast			
Country	airline	ss	ds	overall	airline	ss	ds	overall
Austria	level	log	log	ds+log	log	log*	level	ss+log
Cyprus	level	level	level	ds+level	level	level	level*	ds+level
Czech Rep.	level	level	log	ds+log	level*	level	level	airline+level
Denmark	level	level	level	airline+level	level	level	level	ds+level
Finland	log	level	level	ss+level	level*	log	log*	ds+log
France	level	level	log	airline+level	level	level	level	ds+level
Germany	level	level	log	airline+level	level	log	log	ds+log
Greece	log	level	level	ds+level	log	level	level	ss+level
Hungary	level	log	level	airline+level	level	level	log	ss+level
Ireland	log	level*	level	ds+level	log	level	level	ds+level
Lithuania	log	level	log	ss+level	log	log*	log*	airline+log
Malta	log	level	log	airline+log	log	level	level	ss+level
Netherlands	log	log	level	ss+log	level	level	level*	airline+level
Poland	level	log	level*	ds+level	level	log	log	ds+log
Portugal	level	log	level	ds+level	level	level	level	ds+level
Slovakia	level*	log	level	airline+level	level	log	level	ds+level
Slovenia	level	log	log	airline+level	level	level	log	ss+level
Sweden	level	log	level	airline+level	level	log	level	airline+level
UK	log	level	level	airline+log	log	level	level	airline+log
USA	log	level	level	airline+log	log	level	log*	airline+log
Belgium	log	log	level	airline+log	level	log	log	ss+log
Italy	level	level	level	airline+level	level	level*	level*	ss+level
Luxembourg	level	log	log	airline+level	log	level	log	ds+log
Spain	log	log	log	ds+log	level	log*	log	ss+log

Note: ss - stochastic seasonality subset AR model, ds - deterministic seasonality subset AR model. The underlying RMSEs are based on 24 forecasts. An asterisk (*) indicates a significant improvement according to a 5% level modified Diebold-Mariano test as proposed by Harvey et al. (1997).

Table 6: Forecasting Models for Total CPI Series for Sample Period 1996M1-2007M12, Lag Order Selection Based on AIC

	best 1-step forecast				best 12-step forecast			
Country	airline	ss	ds	overall	airline	ss	ds	overall
Austria	level	log	log	airline+level	log	log	log	airline+log
Cyprus	level	level	log	ds+log	level	level*	level*	ds+level
Czech Rep.	level*	level	level	ds+level	level*	log	level	ds+level
Denmark	level	level	log	ds+log	log	level	level	ds+level
Finland	level	log	level	ss+log	level	level	level	ds+level
France	level	level	log	airline+level	log	level*	log	ds+log
Germany	level	level*	log	airline+level	log	log	log	airline+log
Greece	level	log	level	ss+log	level	level	level	ds+level
Hungary	level	log	level	ds+level	level	level*	level	ds+level
Ireland	level	log	level	ds+level	level	level	level	ds+level
Lithuania	level	level*	level	ds+level	level*	log	log	ds+log
Malta	log	level	level	ss+level	log	level	level	ss+level
Netherlands	log	level	level	ds+level	log	level*	level*	airline+log
Poland	level	log	level*	ss+log	level*	level	level*	ss+level
Portugal	log	log	level	airline+log	log	log	log	ds+log
Slovakia	level	log	level	ds+level	level*	level*	level*	ss+level
Slovenia	level	log	level	ss+log	level	level	log	ds+log
Sweden	level	log	level	ds+level	level	log*	log	ds+log
UK	log	log	log	airline+log	log*	log	log*	airline+log
USA	log	log	log	airline+log	log	log*	log	ds+log
Belgium	level	log	log	airline+level	level	log	log	ds+log
Italy	log	level	log	airline+log	level	level	log	ds+log
Luxembourg	log	log	log	airline+log	log	log	log	ds+log
Spain	log	log	log*	ds+log	level	log*	log*	ds+log

Note: ss - stochastic seasonality full AR model, ds - deterministic seasonality full AR model. The underlying RMSEs are based on 24 forecasts. An asterisk (*) indicates a significant improvement according to a 5% level modified Diebold-Mariano test as proposed by Harvey et al. (1997).

Table 7: Forecasting Models for Total CPI Series for Sample Period 1996M1-2007M12, Lag Order Selection Based on SC

	best 1-step forecast				best 12-step forecast			
Country	airline	ss	ds	overall	airline	ss	ds	overall
Austria	level	level	log	ds+log	log	log	log	airline+log
Cyprus	level	level	log	ds+log	level	level*	level	ds+level
Czech Rep.	level*	log	level*	ds+level	level*	level	level	ds+level
Denmark	level	level	log	ds+log	log	level	level	ds+level
Finland	level	level	level	ss+level	level	log	log	ds+log
France	level	level	log	ds+log	log	log	log	ds+log
Germany	level	level	log	airline+level	log	log	log	airline+log
Greece	level	log	level	ss+log	level	level	level*	ds+level
Hungary	level	level	level	ds+level	level	level	level	ds+level
Ireland	level	level	log	ds+log	level	log	level	ds+level
Lithuania	level	level*	log	ds+log	level*	log	log	ds+log
Malta	log	level	log	airline+log	log	level	level	ds+level
Netherlands	log	level	level	ds+level	log	level	level*	ss+level
Poland	level	level	level*	ds+level	level*	level	level*	ss+level
Portugal	log	log	level	ss+log	log	log	log	ds+log
Slovakia	level	level	level	ds+level	level*	level	level*	ss+level
Slovenia	level	level	level	ds+level	level	level	level*	ss+level
Sweden	level	log	level	ss+log	level	log	log	ds+log
UK	log	level	log	airline+log	log*	log*	log*	airline+log
USA	log	log	log	airline+log	log	log*	log	ds+log
Belgium	level	log	log	ss+log	level	log	log	ss+log
Italy	log	log	level	airline+log	level	log	log	ds+log
Luxembourg	log	level	log	ds+log	log	log	log	ds+log
Spain	log	log	log	ds+log	level	log	log	ss+log

Note: ss - stochastic seasonality full AR model, ds - deterministic seasonality full AR model. The underlying RMSEs are based on 24 forecasts. An asterisk (*) indicates a significant improvement according to a 5% level modified Diebold-Mariano test as proposed by Harvey et al. (1997).

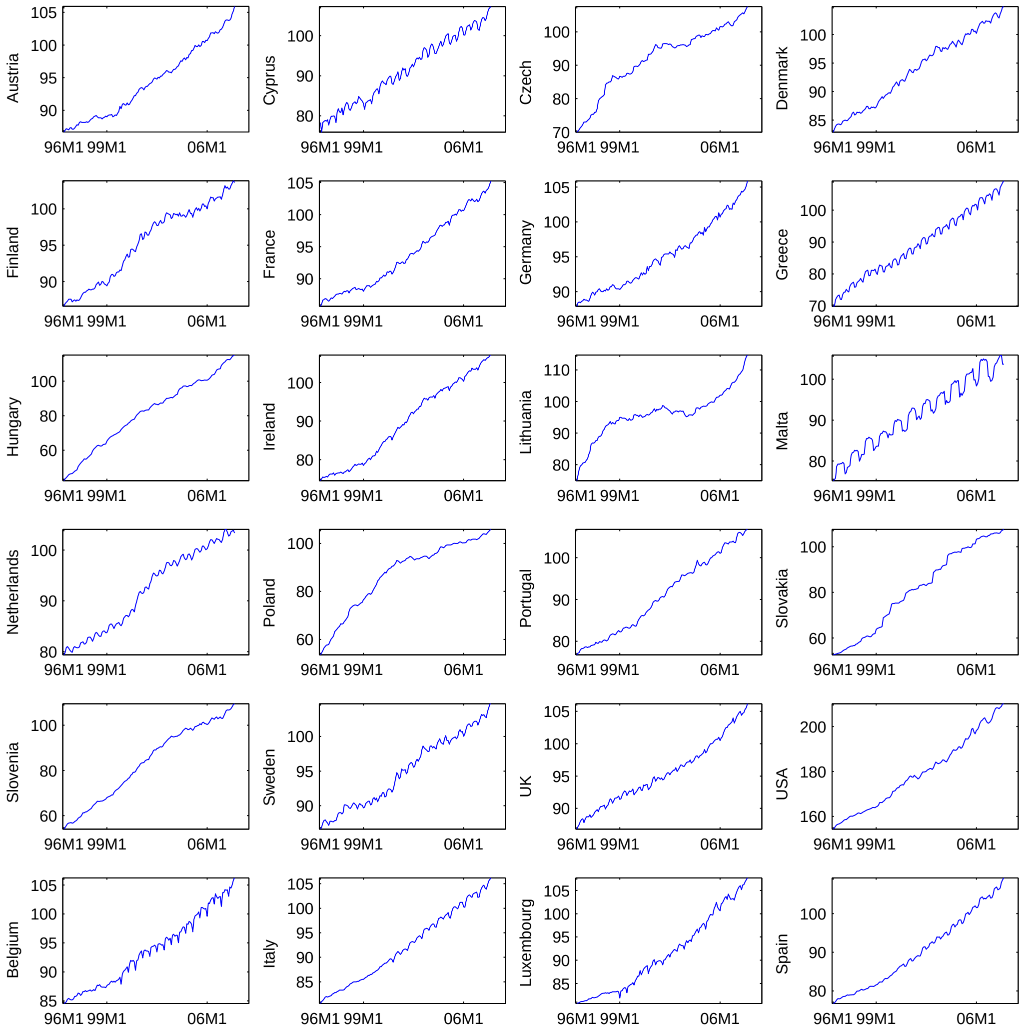

Figure 1: Monthly total CPI series for 1996M1-2007M12.

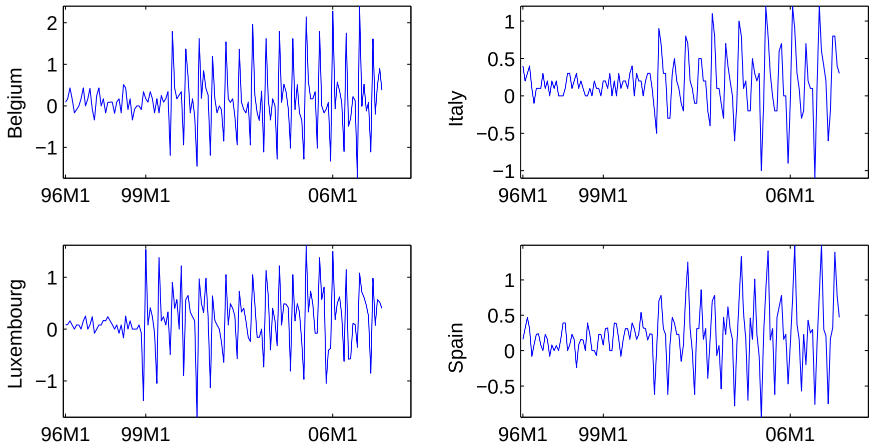

Figure 2: First differences of total CPI series of Belgium, Italy, Luxembourg and Spain for 1996M1-2007M12.

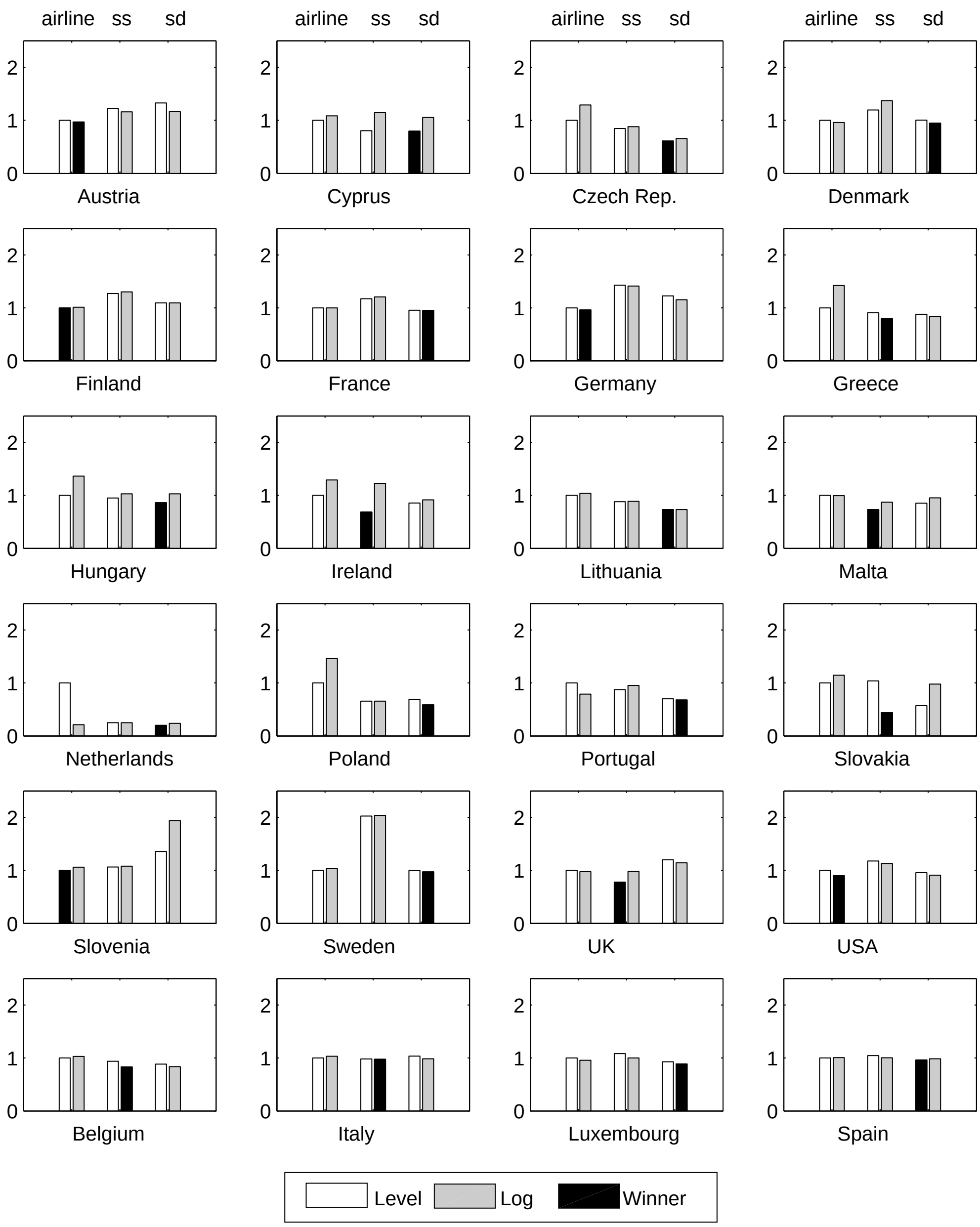

Figure 3: Twelve-step ahead forecast RMSEs relative to $\Delta_s y^{air}_{t+h|t}/y_{t+h-s}$ RMSE for total CPI series based on 24 forecasts; total sample period 1996M1-2007M12; airline - airline model, ss - stochastic seasonality subset AR model, sd - deterministic seasonality subset AR model.

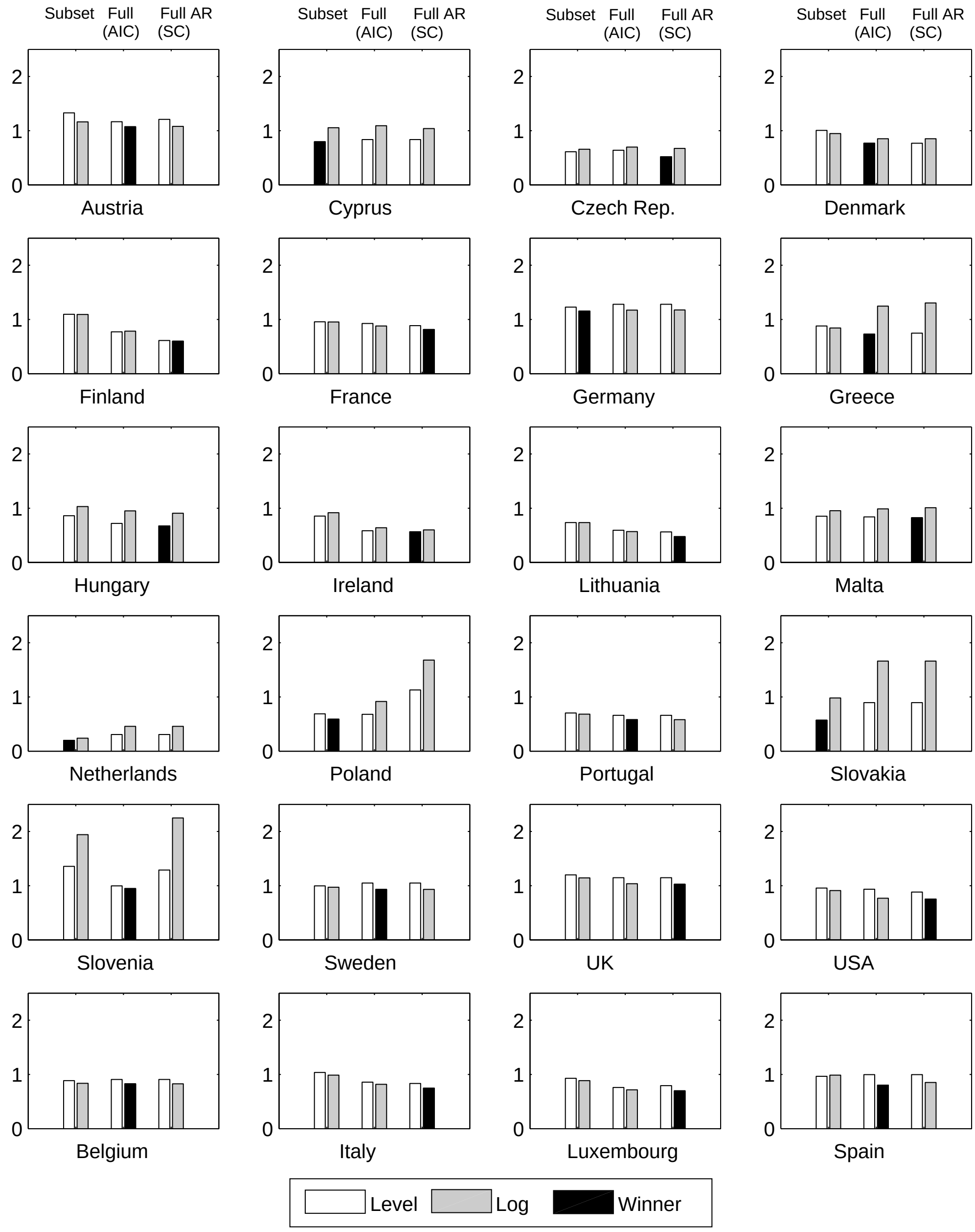

Figure 4: Comparison of twelve-step ahead forecast RMSEs based on subset and full AR deterministic seasonality models for total CPI series based on 24 forecasts; RMSEs relative to RMSE of $\Delta_s y^{air}_{t+h|t}/y_{t+h-s}$; total sample period 1996M1-2007M12.