BANCO DE PORTUGAL

Economic Research Department

FORECASTING INFLATION THROUGH A BOTTOM-UP APPROACH: THE PORTUGUESE CASE

Cláudia Duarte
António Rua

WP 2-05 March 2005

Please address correspondence to Cláudia Duarte, Economic Research Department, Banco de Portugal, Av. Almirante Reis no. 71, 1150-012 Lisboa, Portugal;
Tel: 351–21–3130934 Fax: 351–21–3107804; email: cfduarte@bportugal.pt or António Rua, Economic Research Department, Banco de Portugal, Av. Almirante Reis no. 71, 1150-012 Lisboa, Portugal; Tel: 351–21–3130841 Fax: 351–21–3107804; email: arua@bportugal.pt.

Forecasting inflation through a bottom-up approach: the Portuguese case*

Cláudia Duarte
ISEG

António Rua
Banco de Portugal

Abstract

The aim of this paper is to assess inflation forecasting acurracy over the short-term horizon using Consumer Price Index (CPI) disaggregated data. That is, aggregating forecasts is compared with aggregate forecasting. In particular, three questions are addressed: *i)* one should bottom-up or not, *ii)* how bottom one should go and *iii)* how one should model at the bottom. In contrast with the literature, different levels of data disaggregation are allowed, namely a higher disaggregation level than the one considered up to now. Moreover, both univariate and multivariate models are considered, such as SARIMA and SARIMAX models with dynamic common factors. An out-of-sample forecast comparison (up to twelve months ahead) is done using Portuguese CPI dataset. Aggregating the forecasts seems to be better than aggregate forecasting up to a five-months ahead horizon. Moreover, this improvement increases with the disaggregation level and the multivariate modelling outperforms the univariate one in the very short-run.

Keywords: Inflation forecasting; Bottom-up; SARIMA; SARIMAX; Dynamic common factors.

JEL classification: C22, C32, C43, C53, E31, E37

*We would like to thank Teresa Nascimento for helpful support on the dataset. We would also like to thank Maximiano Pinheiro for helpful comments and suggestions.

1 Introduction

Inflation plays a central role in the economic performance of any country. It is widely accepted that inflation should be neither too high, nor too low. High inflation is seen as an obstacle to economic growth and low (near zero) inflation is associated with the threat of deflation. So, in many countries, and especially in the European Union, the primary objective of the monetary policy is price stability.

According to the European Central Bank (ECB), price stability is defined as "*a year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below*" but "*close to 2% over the medium term*" (ECB (2003a)). Therefore, to ensure that this objective is attained, the monetary authority needs to be constantly monitoring and forecasting the evolution of prices. The existence of lags, caused by transmission mechanisms, and economic shocks, which endanger price stability, explains why inflation forecasting is regarded as a crucial tool for conducting monetary policy. Actually, Jean-Claude Trichet (ECB (2003b)) said that inflation forecasts are "*useful, even indispensable, ingredients of monetary policy strategy*".

Thus, forecasting euro area inflation is very important for monetary policy purposes. However, it is also relevant to forecast country level inflation. First of all, country level inflation forecasting contributes to a better understanding of the different transmission mechanisms in each country. Furthermore, Marcellino, Stock and Watson (2003) found evidence that forecasting inflation at the country level and then aggregating the forecasts increases accuracy against forecasting at the aggregate level. Finally, the usefulness of inflation forecasts is not restricted to monetary policy purposes. Assessing inflation forecasts is also quite relevant in other areas, such as fiscal policy, wage bargaining and financial markets.

One possible way of improving forecast accuracy is by considering more data, in particular, disaggregated one. Some studies have focused on whether using this kind of data increases forecasting accuracy. If it does, this would mean that aggregating the forecasts of disaggregated series would be better than forecasting the aggregate directly. For example, Lütkepohl (1984) says that, "*if the disaggregated data are generated by a known vector ARMA process, it is preferable to forecast the disaggregated variables first and then aggregate the forecasts, rather than forecast the aggregated time series directly*". However, in practice, this is not always true, because of parameter and model uncertainty. The author presents evidence that suggests "*that the forecasts from the aggregated process will be superior to the aggregated forecasts from the disaggregated process for large lead times h if the orders of the processes are unknown*". So, does contemporaneous aggregation of disaggregated forecasts improve forecasting accuracy? The answer to this question is not clear-cut. One advantage of the bottom-up approach is the possibility of capturing idiosyncratic characteristics of each variable by modelling each one individually. However, disaggregated forecast inaccuracy might increase if models are misspecified. Also, what happens with forecast errors is not unambiguous. Forecast errors of the disaggregated variables might cancel out or not.

The aim of this paper is threefold. First, we try to assess if forecasting consumer price index (CPI) subcomponents individually and then aggregating those forecasts (indirect approach) is better than forecasting the aggregate index (direct approach). Currently, there seems to be some evidence in favour of the bottom-up approach for short-term inflation forecasting. For example, Hubrich (2003) and Benalal *et al.* (2004) conclude that, for the euro area, the bottom-up approach is relevant in the very short-term while Fritzer, Moser and Scharler (2002) and Reijer and Vlaar (2003) found that it is also important up to six-months ahead (for Austria and Netherlands, respectively). However, Espasa, Poncela and Senra (2002) found that, for the US, CPI disaggregated forecasting only improves accuracy from the four-months ahead forecast horizon onwards (Espasa, Senra and Albacete (2001) obtained similar results for the euro area).

Additionally, we consider different levels of CPI disaggregation for the bottom-up approach. The above-mentioned papers use a rather low level of disaggregation. In general, the aggregate index is divided in five components namely, unprocessed food, processed food, non-energy industrial goods, energy and services. It is quite reasonable to believe that the results would not remain unchanged if other levels of disaggregation are considered. This paper tries to provide further insight into this question, by considering three different CPI disaggregation levels: the lowest disaggregation level, given by the aggregate price index itself; an intermediate level, in which appear the traditional five components; and a higher disaggregation level, with 59 subcomponents.

Finally, modelling is also important for the bottom-up approach results. For example, Hubrich (2003) found that VAR models dominate simple AR models, while Espasa, Poncela and Senra (2002) conclude that ARIMA models outperform VECM and dynamic factor models. Fritzer, Moser and Scharler (2002) also found that ARIMA models improve on VAR models for shorter horizons (up to six-months ahead). Therefore, both univariate and multivariate models are considered namely, random walks (RW), Seasonal Autoregressive Integrated Moving Average (SARIMA) models, and SARIMA models including exogenous variables (SARIMAX or transfer function models).

The RW model is an obvious benchmark, the SARIMA model tries to capture the variable dynamics based on its past behaviour and the SARIMAX model allows for additional input variables. In particular, the exogenous variables one uses are the common dynamic factors, extracted from the large disaggregated dataset, following Stock and Watson (1998). The purpose of such common factors is to account for potentially relevant information about the variables co-movements, since VAR approach reveals to be intractable when working with 59 variables.

The forecasting performance of the different approaches and models is evaluated by an out-of-sample forecast exercise. The criterion used to compare the forecasting performance of the different methods is the root mean squared forecast error (RMSFE). To test whether the differences are statistically significant or not we use Diebold and Mariano (1995) test.

The results obtained are for the Portuguese case. We find that aggregating the forecasts seems to be significantly better than the aggregate forecasting up to a five-months ahead horizon. Furthermore, the gain of aggregating forecasts against aggregate forecasting is higher when disaggregation level increases and the multivariate modelling outperforms the univariate one in the very short-run.

The remainder of the paper is organised as follows. In section 2, a description of data is given. In section 3, modelling is discussed and in section 4, inflation forecasts accuracy is evaluated. Finally, section 5 concludes.

2 Data

The dataset refers to Portuguese CPI and covers the period from January 1988 to December 2004, comprising 204 observations. During this period, Portuguese CPI suffered several changes[1]. Therefore, the subcomponents used in this paper result from a conciliation effort of the various indices available at each moment. The series were chained with month on month growth rates.

We exclude from our analysis administered and housing prices. In the first case, administered prices behaviour is hardly captured by an econometric model since these prices are adjusted according to specific national regulations[2]. The reasons that justify the exclusion of housing prices are of a different kind. Before 1997, housing price series were collected on an annual frequency only. Hence, this prevents us of including these prices on our monthly dataset.

Thus, we use monthly data for aggregate Portuguese CPI, its partition in five components and in 59 subcomponents. Each one corresponds to a different aggregation level. From the highest to the lowest, we begin with the aggregate index itself. Then, we have the set comprising five product categories (unprocessed food, processed food, non-energy industrial goods, energy and services), which corresponds to an intermediate aggregation level. Finally, the most disaggregated dataset includes 59 subcomponents (see table 1).

Prior to modelling, data are transformed and examined to account for pos-

[1] In particular, four different basis (1983, 1991, 1997 and 2002) as well as methodological changes.

[2] We also exclude fuel prices because they were subject to regulation until quite recently.

sible factors that can distort future analysis. First, all series are transformed to logarithms. Second, following Marcellino, Stock and Watson (2003), it was not found evidence of the presence of large outliers.

3 Model selection

3.1 Preliminary issues[3]

As Diebold and Kilian (2000) point out, unit root pre-testing can be very useful for model selection purposes. They found "*strong evidence that pre-testing improves forecast accuracy relative to routinely differencing the data*". So, in order to check the order of integration of the variables, unit root tests are carried out.

Before performing the tests, it is useful to analyse the graphics of the original series (in natural logarithms) and its first difference. In general, the logarithms of price indices reveal a smooth upward trend and its first difference shows an erratic behaviour around a constant. Accordingly, the price indices, previously transformed to logarithms, appear to be integrated of order one.

Additionally, we perform three different kinds of unit root tests. In first place, Dickey and Pantula (1987) tests are carried out. These authors suggested an appropriate sequence of tests to handle situations in which the order of integration is higher than one. These tests indicate that price indices are not integrated of order two (I(2)) but are integrated of order one (I(1)). The latter evidence is also supported by Augmented Dickey-Fuller (ADF) tests. Kwiatkowski, Phillips, Schmidt and Shin (1992) proposed an alternative test, known as KPSS test. This test rejects the null hypothesis of stationarity for the levels. Thus, it seems that price indices are I(1). Among others, Hubrich (2003), Fritzer, Moser and Scharler (2002) and Meyler, Kenny and Quinn (1998) obtain similar results.

Following Hylleberg *et al.* (1990), seasonal unit root tests are also carried out. In particular, it is used the test procedure for monthly data (see Beaulieu

[3] The results discussed in this section are available from the authors upon request.

and Miron (1993)). Since our interest relies on testing the presence of stochastic integrated seasonality, we include seasonal dummies in test equations to control for deterministic seasonality. In general, we reject the null hypothesis, which means that we reject unit roots at most frequencies. Based on this, there seems to be no reason for seasonal differencing.

3.2 Modelling

Forecasting results can be affected by several factors. Two of those factors are the type of models chosen and the model selection criteria. There are two main types of models - causal and non-causal models. Traditionally, causal information is seen as more important than non-causal. Causal models tend to have smaller forecast errors than the non-causal ones. However, sometimes, simple time-series models perform better than structural models. For example, according to Hendry (2002), simpler models deliver better results because they are more robust to structural breaks. Hendry points out that "*unless the model coincides with the generating mechanism, one cannot even prove that causal variables will dominate non-causal in forecasting*". Moreover, univariate models can be a reasonable, or even better, alternative to more complex models, in particular, for short-term forecasting (see, for example, Fildes and Stekler (1999)).

Regarding model selection, there are two main criteria - in-sample and out-of-sample methods. Frequently, their results are mixed. Models with the best in-sample fit are not necessarily the best forecasting models. However, selecting a model based on its out-of-sample performance means that its selection will strongly rely on a short sample period. According to Hendry and Clements (2001), "*forecasting success is not a good index for model selection*" and "*forecast failure is equally not a ground for model rejection (...). Consequently, a focus on 'out-of-sample' forecast performance to judge models (...) is unsustainable*". Moreover, Inoue and Kilian (2003) show that, under standard conditions, the in-sample method is more reliable than the out-of-sample one. Therefore, we focus on in-sample analysis for model selection.

3.2.1 Univariate modelling

We first proceed into univariate modelling. In particular, SARIMA models are considered. Although SARIMA models are based on the series past behaviour only, one should note that these models are able to capture rich dynamics, both seasonal and non-seasonal.

The SARIMA modelling follows Box and Jenkins (1976) methodology. This methodology comprises three stages: identification, estimation and diagnostic checking. In the first stage, one begins by plotting the autocorrelation (ACF) and partial autocorrelation (PACF) functions. The corresponding visual inspection gives a first idea of the order of integration of the variables. An autocorrelation function dying out slowly suggests that the series is non-stationary. Nowadays, this can be complemented with the above-mentioned unit root tests for a more formal procedure. Non-stationary variables are transformed to become stationary. Afterwards, one needs to infer the form of the SARIMA model. Thus, ACF and PACF plots of the transformed series are examined. These plots reflect patterns that suggest appropriate orders of the autoregressive and moving average polynomials. In the second stage, the model chosen in the previous stage is estimated.

Finally, in the last stage, the estimated model is evaluated according to several criteria. Among others, information criteria (like, for example, Akaike Information Criterion (AIC) or Schwarz Bayesian Criterion (SBC)) are very useful goodness-of-fit measures, especially because they account for parsimony. Box and Jenkins argue that parsimonious models can be more reliable for forecasting than overfitted models. Moreover, it is also useful to plot the residuals to look for outliers as well as testing for serial autocorrelation in the residuals. This process is iterative, that is, when the model chosen is not satisfactory, a new cycle begins and the same steps are repeated until a suitable model is found.

In our case, we first difference all series, since unit root tests and ACF plots indicated that price indices are I(1). Likewise, no seasonal differencing is done. Whenever it seems appropriate, seasonal dummies are added to the models. After testing for stochastic integrated seasonality, stochastic stationary seasonality is modelled through the seasonal polynomials while seasonal dummy variables account for deterministic seasonality. The model for series y has the following form:

$$\phi(L)\varphi(L^s)(\Delta y_t - \alpha - \beta_1 D_1 - ... - \beta_{11} D_{11}) = \theta(L)\delta(L^s)\varepsilon_t \tag{1}$$

where α is a constant, D_i is a seasonal dummy $(i = 1, ..., 11)$, β_i its corresponding coefficient $(i = 1, ..., 11)$ and ε_t is a white noise. The lag polynomials ($\phi(L)$ - autoregressive polynomial; $\varphi(L^s)$ - seasonal autoregressive polynomial; $\theta(L)$ - moving average polynomial; $\delta(L^s)$ - seasonal moving average polynomial) are defined as usual.

The sample used for model estimation runs from January 1988 to December 2000. The SARIMA models are estimated by non-linear least squares. These models are selected resorting to coefficient significance tests, SBC, residual correlation plots and Ljung-Box tests. Thus, 65 models (59 subcomponents plus 5 components and plus the aggregate index itself) were chosen (see table 2).

In contrast with some literature (see, for example, Stock and Watson (1999) or Marcellino, Stock and Watson, (2003)), models are not specified as a linear projection of the h-step ahead (h = 1,. . . ,12) interest variable onto t-dated and lagged regressors. The latter is called 'direct forecasting' while here we follow 'iterated forecasting'. 'Iterated forecasting' is done by using a one-period model iterated forward. In fact, Marcellino, Stock and Watson (2004) found that iterated forecasts typically outperform direct forecasts and iterated forecast accuracy increases with the forecast horizon.

3.2.2 Multivariate modelling

When one estimates SARIMA models for each series one is ignoring potentially relevant information about the variables co-movements. Using an alternative time-series technique, namely VAR models, can be seen as a possible solution to this problem. As Granger and Yoon (2001) put it, "*VAR models are the major* *tools for investigating linear relationships between **small** groups of variables*" (our emphasized).

However, this paper considers a large set of variables, rendering VAR models intractable. Therefore, we decided to estimate SARIMAX (or transfer function) models. These models can be seen as hybrid models: they are not the typical causal (structural) models, but they are certainly not univariate models. Nevertheless, they are very appealing because they allow one to extend the univariate models by including exogenous variables that affect the dynamic behaviour of the dependent variable.

The additional variables considered are the dynamic common factors, extracted from the large disaggregated dataset comprising the 59 subcomponents. The key role of the common factors is to summarise large amounts of information in a few variables, which capture the main features of the original data. In fact, the idea behind the factor model is that variables have two components: the common component, which can be captured by a small number of variables – the common factors; and the idiosyncratic component, which reflects variable-specific features. Hence, the purpose of using common factors is to reduce the dimension of data, by pooling the most significant information from the initial series while excluding their idiosyncratic component.

For the dynamic common factors extraction, we follow Stock and Watson (1998). According to them, it is possible to estimate dynamic common factors consistently in an approximate dynamic factor model, when both time series and cross-sectional dimensions are large.

Let X_t be a N-dimensional multiple time series of variables, observed for $t = 1, \ldots, T$. Assume that the dynamic factor model can be represented by

$$X_{it} = \lambda_i(L) f_t + e_{it} \tag{2}$$

for $i = 1, \ldots, N$, where e_t is the $(N \times 1)$ idiosyncratic disturbances vector, $\lambda_i(L)$ are lag polynomials in nonnegative powers of L and f_t is the r^* common dynamic factors vector. If we assume that $\lambda_i(L)$ have finite lags (for example, m lags), then it is possible to rewrite the dynamic factor model with time invariant parameters. Thus, the model can be redefined as

$$X_t = \Lambda F_t + e_t \tag{3}$$

where $F_t = (f'_t, \ldots, f'_{t-m})'$ is a $r = r^*(m+1)$ vector of stacked vectors and Λ is a $(N \times r)$ parameter matrix. The main advantage of the approximate dynamic factor model is that it can be consistently estimated by principal components.

Due to different model representations, factors can be extracted alternatively from the contemporaneous values of X_t only, or from a stacked set of variables, including X_t and its lagged values (see Stock and Watson (1998)). Theoretically, adding more variables (lagged values of X_t) could lead to an improvement in the finite sample performance of the models. However, Stock and Watson (2002) conclude that, for US monthly price series, "*forecasts based on the stacked data perform less well than those based on the unstacked data*". Therefore, as Angelini, Henry and Mestre (2001a, 2001b) and Marcellino, Stock and Watson (2003), we extract the common factors from the contemporaneous values of X_t only.

Before extracting the factors, the price series are subject to preliminary transformations. First of all, the 59 subcomponents are transformed to logarithms and first differenced. Afterwards, since we are not interested in capturing spurious relations based on common seasonal patterns, the series are seasonally adjusted[4]. In fact, what we want to capture is the underlying non-seasonal co-movement of prices. Finally, all series are standardized.

Once obtained the common factors, the next step is to consider them in modelling. In particular, using common factors as exogenous variables in the SARIMAX models has some advantages. There are gains in terms of additional information that is brought into the analysis (especially, the one about variables co-movement) and the number of variables in the model does not increase substantially. Moreover, Stock and Watson (1998) show that the estimated factors can efficiently replace the true factors in forecasting models.

[4] Resorting to X-12 ARIMA seasonal adjustment procedure.

One should note that the transfer function models are based on the assumption of input variables exogeneity. By definition, the common factors are not exogenous variables that evolve independently from the price series, from which they are extracted. However, the common factors are linear combinations of all those 59 price series. In the whole, the weight associated to each original variable is rather small. Therefore, it is assumed that the common factor is an exogenous variable, neglecting the quantitatively small effect of contemporaneous correlation that possibly exists between the common factor and each price series innovations.

A key issue is the determination of the number of factors to include in the model. Bai and Ng (2002) developed criteria for that purpose. These criteria are similar to the well-known information criteria (AIC and SBC, among others) but the penalty is also a function of the cross-sectional dimension (N). These criteria are valid for the approximate dynamic factor model only. The criteria are[5]:

$$IC_1(r) = \ln(V(r,F)) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)$$

$$IC_2(r) = \ln(V(r,F)) + r\left(\frac{N+T}{NT}\right)\ln C_{NT}^2$$

$$IC_3(r) = \ln(V(r,F)) + r\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right)$$

where $V(r,F) = (1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \Lambda_i F_t)^2$ and C_{NT}^2 is the minimum of $\left\{\sqrt{N}, \sqrt{T}\right\}$. The results presented by Bai and Ng suggest that IC_3 is less robust than IC_1 and IC_2. In practice, one must arbitrarily choose an r_{max} for starting the calculations. The optimal number of factors minimizes the information criteria.

The results suggest the relevance of one factor only (the first principal component) (see table 3). This evidence is also supported by the rule-of-thumb

[5] Here, we only present three of the six criteria developed by Bai and Ng (2002). The other three criteria have a different formulation, but deliver equivalent results.

based on the variance of the original series explained by each factor. As it can be seen in figure 1, there is a significant difference between the variance explained by the first principal component and by the other principal components, which also suggests the truncation in favour of just one factor.

The SARIMAX model considered for each series y can be written as

$$\phi(L)\varphi(L^s)(\Delta y_t - \alpha - \beta_1 D_1 - ... - \beta_{11} D_{11} - \upsilon(L)x_t) = \theta(L)\delta(L^s)\varepsilon_t \qquad (4)$$

where α is a constant, D_i is a seasonal dummy ($i = 1, ..., 11$), β_i its corresponding coefficient ($i = 1, ..., 11$) and ε_t is a white noise. The lag polynomials ($\phi(L)$ - autoregressive polynomial; $\varphi(L^s)$ - seasonal autoregressive polynomial; $\theta(L)$ - moving average polynomial; $\delta(L^s)$ - seasonal moving average polynomial; $\upsilon(L)$ - polynomial associated with the exogenous variable x_t) are defined as usual.

The SARIMAX modelling also follows Box and Jenkins (1976). The identification comprises five stages[6]. The first one consists in fitting an ARMA model to the exogenous variable, that is, the common factor. Following the univariate modelling strategy, an AR(1) model was chosen. The corresponding residuals are the filtered values of the exogenous variable. By applying the same filter to the variable of interest, in the second step, we obtain the filtered values of the price series. In the next step, both filtered series are used to build a cross-correlogram. The pattern exhibited by the cross-correlations between the common factor and price series helps to determine the number of lags of both variables that should be introduced in the SARIMAX models.

The fourth step consists in estimating plausible models of the following form

$$\phi(L)\varphi(L^s)(\Delta y_t - \alpha - \beta_1 D_1 - ... - \beta_{11} D_{11} - \upsilon(L)x_t) = e_t \qquad (5)$$

and selecting the model with the best fit. The residuals of the resulting model (e_t) are not necessarily white noise. So, the examination of the residual

[6]See, for example, Enders (2004) for more details.

autocorrelation should suggest plausible orders for the $\theta(L)$ and $\delta(L^s)$ polynomials. The last step consists in estimating altogether the SARIMAX model.

Likewise the univariate case, 65 SARIMAX models are selected (see table 4). Note that, we also introduced the common factor in the five components and aggregate index equations. This allows us to compare both the gains stemming from the disaggregated dataset as well as from the forecasting methods.

4 Out-of-sample forecast evaluation

Forecasting performance is evaluated through an out-of-sample forecast exercise. For each series and model, a recursive estimation process is implemented. Starting from the estimation period (January 1988 to December 2000), each round a new observation is added to the sample. The last sample considered runs from January 1988 to December 2003. In each round of this recursive estimation process one to twelve step ahead forecasts are computed. Thus, for each forecast horizon, 37 observations are available.

For each forecast horizon, the forecast series of all 59 subcomponents are aggregated, using the corresponding weights of Portuguese CPI. Then, from the index forecasts, year-on-year inflation rate forecasts are obtained. The same is done for the intermediate disaggregation level (5 components). The forecasts that result from aggregating forecasts are called ‘indirect’. ‘Direct’ forecasts are obtained by forecasting the aggregate index. Likewise, year-on-year inflation rate forecasts are computed.

The forecasting performance is evaluated by comparing RMSFE. The RMSFE is one of several statistics, and definitely the most used, which can be calculated to assess the performance of out-of-sample forecasts. Fritzer, Moser and Scharler (2002) argue that the RMSFE is particularly suitable in this context because the implicit central bank’s loss function related to inflation deviations appears to be quadratic or some transformation of it. However, the RMSFE does not tell us anything about the significance of the differences between forecasts of competing models. Diebold and Mariano (1995) proposed a test to evaluate the significance of the difference between forecasts obtained by different methods. Consider two models (A and B), which produce forecasts for the variable y. Let $\{\hat{y}_{At}\}_{t=1}^{T}$ and $\{\hat{y}_{Bt}\}_{t=1}^{T}$ be the corresponding sequences of forecasts. The associated forecast error sequences are denoted by $\{e_{At}\}_{t=1}^{T}$ and $\{e_{Bt}\}_{t=1}^{T}$. Since larger errors mean less accurate forecasts, it is possible to build a loss function associated with those errors, $g(e_{it})$ with $i = A, B$. Most commonly, quadratic loss functions are chosen. The Diebold-Mariano test is based on the difference of the loss functions for both kinds of forecasts, that is, $d_t = g(e_{At}) - g(e_{Bt})$. Under the null hypothesis, forecasts are equally accurate, so $d_t = 0$.

Even though the forecast evaluation framework is similar for univariate and multivariate models, multivariate forecasting requires some additional steps. The reason why this happens is quite obvious. Forecasting the dependent variable also requires forecasting the exogenous variable.

However, forecasting the common factor is not a straightforward issue. It is possible to obtain 'direct' or 'indirect' factor forecasts. The direct forecasts can be obtained by fitting a model to the common factor and by using it to produce one to twelve months ahead forecasts, for each recursive sample[7]. In particular, we use the model already fitted to the common factor in the previous section. Obviously, this brings up another drawback - the potential misspecification of the common factor model. Then, these factor forecasts are used as inputs in the SARIMAX models for forecasting purposes.

The 'indirect' approach relies on the fact that the common factor is extracted from the price series. Therefore, implicit common factor forecasts can be obtained from price series forecasts. For each sample, the observed price series are stacked with one to twelve months ahead forecasts, resulting from the univariate models. Afterwards, the common factor is extracted from these enlarged series and used for SARIMAX forecasting.

The results obtained are quite interesting. Unsurprisingly, the RW models are the ones that present the worst performance (see figures 2, 3 and 4). For the

[7] A fully recursive methodology is applied. This means that for each sample the series are seasonally adjusted and standardized, and the corresponding common factor is extracted.

RW, considering the intermediate disaggregation level does not improve forecasts accuracy against direct forecasting (see figure 5 and table 6). However, considering the highest disaggregation level (59 subcomponents) improves forecasts accuracy for 3 to 5 step ahead forecast horizon but this improvement is not statistically significant (see table 7).

Regarding SARIMA models, for forecasting horizons up to five months ahead, the indirect forecast does better than the direct forecast. Up to four months ahead, this is true for both disaggregation levels (five components and 59 subcomponents) (see figure 6). However, within these forecast horizons, the highest disaggregation level delivers better results than the five components level. Furthermore, for 2 and 3 months ahead forecast horizons, SARIMA indirect forecasts pooled from the 59 subcomponents are statistically better than the corresponding direct forecasts[8] (see table 9). From 6 to 12 months ahead, SARIMA direct forecasts present the lowest RMSFE of all models considered (see table 5). Thus, it seems that, for short-term forecasting, the gain in terms of additional information stemming from disaggregated price data through a bottom-up approach is such that compensates the loss due to potential model misspecification and parameter uncertainty.

Concerning SARIMAX models, a similar picture arises regarding direct vs. indirect forecasting (see figures 7 and 8 and tables 10, 11, 12 and 13). Moreover, multivariate models, which consider more information through dynamic common factors, seem to improve on SARIMA models for the very short-run forecasting. In particular, the performance of SARIMAX models using the AR(1) model for forecasting the common factor which consider the highest disaggregation level is the best for one and two months ahead horizons.

[8] Considering a significance level of 5 per cent. For a 10 per cent significance level, the difference for the four months ahead forecast horizon is also significant.

5 Conclusion

The purpose of this paper is to assess if one can improve forecasting accuracy by considering disaggregated price data. In particular, three issues are addressed: *i)* one should bottom-up or not, *ii)* how bottom one should go and *iii)* how one should model at the bottom.

In contrast with the literature, different levels of data disaggregation are allowed. This paper considers three CPI disaggregation levels: the lowest disaggregation level, given by the aggregate price index itself; an intermediate level, in which appear five components; and a higher disaggregation level, with 59 subcomponents. Furthermore, both univariate and multivariate models are considered, such as SARIMA and SARIMAX models with the dynamic common factors of CPI disaggregated data as exogenous regressors.

The forecasting accuracy (up to twelve months ahead) of the bottom-up approach is evaluated by an out-of-sample forecast exercise using Portuguese CPI dataset. We find that the bottom-up approach seems to improve substantially forecasts accuracy up to a five-months ahead horizon. Moreover, the gain of aggregating forecasts against aggregate forecasting is higher when disaggregation level increases and the multivariate approach outperforms the univariate one in the very short-run, in terms of RMSFE.

References

[1] Angelini, E., Henry, J., Mestre, R. (2001a) Diffusion index-based inflation forecasts for the Euro Area. ECB Working Paper no 61.

[2] Angelini, E., Henry, J., Mestre, R. (2001b) A Multi Country Trend Indicator for Euro Area Inflation: Computation and Properties. ECB Working Paper no 60.

[3] Bai, J., Ng, S. (2002) Determining the Number of Factors in Approximate Factor Models. Econometrica 70(1): 191-221.

[4] Beaulieu, J., Miron, J. (1993) Seasonal unit roots in aggregate US data. Journal of Econometrics 55: 305-328.

[5] Benalal, N., del Hoyo, J., Landau, B., Roma M., Skudelny, F. (2004) To aggregate or not to aggregate? Euro area inflation forecasting. ECB Working Paper no 374.

[6] Box, G., Jenkins, G. (1976) Time Series Analysis: forecasting and control. Holden-Day.

[7] Dickey, D., Pantula, S. (1987) Determining the order or differencing in autoregressive processes. Journal of Business and Economic Statistics 5(4): 18-24.

[8] Diebold, F., Kilian, L. (2000) Unit-Root Tests Are Useful for Selecting Forecasting Models. Journal of Business and Economic Statistics 18(3): 265-273.

[9] Diebold, F., Mariano, R. (1995) Comparing Predictive Accuracy. Journal of Business and Economic Statistics 13(3): 253-263.

[10] Enders, W. (2004) Applied Econometric Time Series, 2^{nd} edition. Wiley.

[11] Espasa, A., Poncela, P., Senra, E. (2002) Forecasting Monthly US Consumer Price Indexes Through a Disaggregated I(2) Analysis. Mimeo.

[12] Espasa, A., Senra, E., Albacete, R. (2001) Forecasting Inflation in the European Monetary Union: A Disaggregated Approach by Countries and by Sectors. Univ. Carlos III de Madrid Working Paper no 37.

[13] European Central Bank (2003a) The ECB's monetary policy strategy. ECB Press Release, 8 May 2003.

[14] European Central Bank (2003b) The ECB's monetary policy strategy after the evaluation and clarification of May 2003. Speech by Jean-Claude Trichet, 20 November 2003.

[15] Fildes, R., Stekler, H. (1999) The State of Macroeconomic Forecasting. Lancaster University Management School Working Paper no 2.

[16] Fritzer, F., Moser, G., Scharler, J. (2002) Forecasting Austrian HICP and Its Components Using VAR and ARIMA Models. Oesterreichische Nationalbank Working Paper no 73.

[17] Granger, C., Yoon, G. (2001) Self-generating variables in a cointegrated VAR framework. University of California San Diego Discussion Paper no 4.

[18] Hendry, D. (2002) Forecast Failure, Expectations Formation and the Lucas Critique. Nuffield College Economics Working Paper no 8.

[19] Hendry, D., Clements, M. (2001) Economic Forecasting: Some Lessons from Recent Research. ECB Working Paper no 82.

[20] Hubrich, K. (2003) Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? ECB Working Paper no 247.

[21] Hylleberg, S., Engle, R, Granger, C, Yoo, B. (1990) Seasonal Integration and Cointegration. Journal of Econometrics 44: 215-238.

[22] Inoue, A., Kilian, L. (2003) On the selection of forecasting models. ECB Working Paper no 214.

[23] Kwiatkowski, D., Phillips, P., Schmidt, P., Shin, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54: 159-178.

[24] Lütkepohl, H. (1984) Forecasting Contemporaneously Aggregated Vector ARMA Processes. Journal of Business and Economic Statistics 2(3): 201-214.

[25] Marcellino, M., Stock, J. H., Watson, M. W. (2003) Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. European Economic Review 47: 1-18.

[26] Marcellino, M., Stock, J. H., Watson, M. W. (2004) A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. Mimeo.

[27] Meyler, A., Kenny, G., Quinn, T. (1998) Forecasting Irish Inflation Using ARIMA Models. Central Bank and Financial Services Authority of Ireland Research Technical Paper no 3.

[28] Reijer, A., Vlaar, P. (2003) Forecasting inflation: An art as well as a science! De Nederlandsche Bank Staff Reports no 107.

[29] Stock, J. H., Watson, M. W. (1998) Diffusion Indexes. NBER Working Paper no 6702.

[30] Stock, J. H., Watson, M. W. (1999) Forecasting inflation. Journal of Monetary Economics 44: 293-335.

[31] Stock, J. H., Watson, M. W. (2002) Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics 20(2): 147-162.

Table 1 - Variables list

	Subcomponents	Components
Series 1	Rice	Pf
Series 2	Other cereal products	Pf
Series 3	Pasta products	Pf
Series 4	Bread and other bakery products	Pf
Series 5	Potatoes and other tubers	Unpf
Series 6	Dried vegetables	Unpf
Series 7	Fresh and preserved vegetables	Unpf
Series 8	Fruit	Unpf
Series 9	Meat of sheep and goat	Unpf
Series 10	Meat of swine	Unpf
Series 11	Meat of bovine animals	Unpf
Series 12	Other meat	Unpf
Series 13	Sausages and preserved meat	Unpf
Series 14	Poultry	Unpf
Series 15	Fresh, chilled or frozen fish	Unpf
Series 16	Fresh, chilled or frozen seafood	Unpf
Series 17	Other preserved or processed fish and seafood and fish and seafood preparations	Unpf
Series 18	Dried, smoked or salted fish and seafood	Unpf
Series 19	Eggs	Pf
Series 20	Milk	Pf
Series 21	Yoghurt, cheese and other milk-based products	Pf
Series 22	Edible oils	Pf
Series 23	Butter, margarine and other fats	Pf
Series 24	Sugar and confectionery	Pf
Series 25	Cocoa and powdered chocolate	Pf
Series 26	Coffee	Pf
Series 27	Tea	Pf
Series 28	Sauces, condiments and salt	Pf
Series 29	Baking powders, preparations and soups	Pf
Series 30	Catering	Serv
Series 31	Alcoholic beverages	Pf
Series 32	Mineral waters, soft drinks and juices	Pf
Series 33	Clothing materials and garments	Neig
Series 34	Dry-cleaning, repair and hire of clothing	Serv
Series 35	Footwear	Neig
Series 36	Repair and hire of footwear and shoe-cleaning services	Neig
Series 37	Gas	Enrgy
Series 38	Heating and cooking appliances, refrigerators, washing machines and similar major household appliances	Neig
Series 39	Furniture, furnishings and carpets	Neig
Series 40	Repair of furniture, furnishings and floor coverings	Serv
Series 41	Household textiles	Neig
Series 42	Glassware, tableware and household utensils and tools	Neig
Series 43	Non-durable household goods	Neig
Series 44	Repair of household appliances	Serv
Series 45	Therapeutic appliances and equipment	Neig
Series 46	Medical, paramedical and hospital services	Serv
Series 47	Motor cars, motor cycles and bicycles and spare parts and accessories	Neig
Series 48	Maintenance and repairs; other services in respect of personal transport equipment	Serv
Series 49	Telephone and telefax equipment	Neig
Series 50	Education	Serv
Series 51	Equipment for the reception, recording and reproduction of sound; other major durables for recreation and culture	Neig
Series 52	Repair of audio-visual, photographic and data processing equipment	Serv
Series 53	Receational and cultural services	Serv
Series 54	Newspapers, books and stationery	Neig
Series 55	Accommodation services	Serv
Series 56	Package holidays	Serv
Series 57	Electrical appliances and products for personal care	Neig
Series 58	Hairdressing salons and personal grooming establishments	Serv
Series 59	Insurance and banking services	Serv

Note: Unpf - Unprocessed food; Pf - Processed food; Neig - Non-energy industrial goods; Enrgy - Energy; Serv - Services.

Table 2 - SARIMA models

	p	d	q	P	D	Q	Seasonal dummies
Series 1	3	1	1	0	0	0	y
Series 2	3	1	1	2	0	0	y
Series 3	3	1	1	0	0	0	n
Series 4	1	1	0	0	0	1	n
Series 5	0	1	1	0	0	1	y
Series 6	0	1	3	||2||	0	0	y
Series 7	1	1	1	1	0	1	n
Series 8	1	1	0	0	0	1	y
Series 9	0	1	0	2	0	2	y
Series 10	0	1	0	0	0	0	y
Series 11	1	1	2	0	0	0	y
Series 12	1	1	1	0	0	0	y
Series 13	||3||	1	0	0	0	0	n
Series 14	0	1	2	||2||	0	1	y
Series 15	0	1	1	1	0	0	y
Series 16	2	1	2	0	0	0	y
Series 17	3	1	0	1	0	0	y
Series 18	3	1	0	0	0	0	y
Series 19	1	1	1	0	0	0	y
Series 20	||4||	1	1	0	0	0	n
Series 21	1	1	1	1	0	1	y
Series 22	2	1	0	0	0	0	n
Series 23	1	1	1	0	0	0	n
Series 24	2	1	0	1	0	0	n
Series 25	2	1	1	0	0	0	y
Series 26	||2,3||	1	2	0	0	0	y
Series 27	0	1	1	1	0	1	y
Series 28	0	1	2	0	0	0	y
Series 29	1	1	0	1	0	1	n
Series 30	||1,3||	1	1	1	0	1	n
Series 31	2	1	0	0	0	0	y
Series 32	1	1	1	0	0	0	y
Series 33	3	1	0	1	0	1	y
Series 34	1	1	1	1	0	1	y
Series 35	3	1	0	1	0	1	y
Series 36	2	1	0	1	0	0	y
Series 37	||2||	1	0	1	0	1	n
Series 38	1	1	1	0	0	0	n
Series 39	0	1	2	1	0	1	n
Series 40	1	1	0	2	0	1	y
Series 41	||3||	1	1	1	0	0	n
Series 42	2	1	0	0	0	0	n
Series 43	2	1	1	0	0	0	y
Series 44	0	1	1	2	0	1	y
Series 45	1	1	0	0	0	0	n
Series 46	3	1	1	1	0	0	y
Series 47	||3||	1	||1,3,4||	0	0	0	n
Series 48	1	1	0	1	0	1	n
Series 49	2	1	0	1	0	0	n
Series 50	0	1	0	0	0	0	y
Series 51	0	1	1	1	0	1	n
Series 52	||2||	1	0	1	0	1	y
Series 53	1	1	0	0	0	0	y
Series 54	1	1	0	1	0	0	n
Series 55	1	1	0	1	0	1	y
Series 56	2	1	0	||2||	0	0	y
Series 57	1	1	0	0	0	0	n
Series 58	0	1	1	2	0	1	y
Series 59	0	1	2	2	0	0	y
Unprocessed food	0	1	1	0	0	0	y
Processed food	2	1	0	1	0	0	n
Non-energy industrial goods	||1,2,5||	1	1	1	0	1	y
Energy	||2||	1	0	1	0	1	n
Services	||2,3,4||	1	0	0	0	1	y
Total	1	1	0	1	0	1	y

Note: p - autoregressive polynomial order; d - integration order; q - moving average polynomial order; P - seasonal autoregressive polynomial order; D - seasonal integration order; Q - seasonal moving average polynomial order; ||L|| = lag of order L; y = yes; n = no.

Table 3 - Bai and Ng criteria

Number of factors	IC_1	IC_2	IC_3
r=1	9.005	9.013	8.987
r=2	9.026	9.041	8.988
r=3	9.049	9.071	8.993
r=4	9.072	9.102	8.998
r=5	9.102	9.140	9.009
r=6	9.133	9.178	9.021
r=7	9.162	9.215	9.032
r=8	9.190	9.250	9.041

Table 4 - SARIMAX models

	p	d	q	P	D	Q	R	Seasonal dummies
Series 1	||2||	1	||3||	0	0	0	0	y
Series 2	3	1	0	1	0	0	0	y
Series 3	||2||	1	0	0	0	0	0	n
Series 4	1	1	0	0	0	1	-	n
Series 5	0	1	1	0	0	0	||0,2||	y
Series 6	0	1	3	||2||	0	0	-	y
Series 7	1	1	1	1	0	1	-	n
Series 8	1	1	||4||	0	0	1	||1||	y
Series 9	0	1	0	2	0	2	-	y
Series 10	0	1	0	0	0	0	-	y
Series 11	1	1	2	0	0	0	-	y
Series 12	1	1	1	0	0	0	-	y
Series 13	||3||	1	0	0	0	0	0	n
Series 14	0	1	2	||2||	0	1	-	y
Series 15	0	1	1	1	0	0	1	y
Series 16	0	1	0	0	0	0	0	y
Series 17	||3||	1	0	1	0	0	0	y
Series 18	3	1	0	0	0	0	-	y
Series 19	1	1	1	0	0	0	-	y
Series 20	0	1	||4||	0	0	0	0	n
Series 21	0	1	||4||	1	0	0	1	y
Series 22	2	1	0	0	0	0	0	n
Series 23	0	1	0	0	0	0	0	n
Series 24	||5||	1	0	1	0	0	||0,1,4||	n
Series 25	1	1	0	0	0	0	0	y
Series 26	||2,3||	1	0	0	0	0	||0,4||	y
Series 27	0	1	1	1	0	1	-	y
Series 28	0	1	2	0	0	0	0	y
Series 29	0	1	0	0	0	0	0	n
Series 30	2	1	0	1	0	1	||1||	n
Series 31	1	1	0	0	0	0	||1||	y
Series 32	1	1	||4||	0	0	0	||4||	y
Series 33	3	1	0	1	0	1	-	y
Series 34	2	1	0	0	0	0	1	y
Series 35	3	1	0	1	0	1	-	y
Series 36	1	1	0	0	0	0	1	y
Series 37	||2||	1	0	1	0	1	-	n
Series 38	1	1	1	0	0	0	-	n
Series 39	0	1	1	0	0	0	0	n
Series 40	2	1	0	1	0	0	0	y
Series 41	||3||	1	1	1	0	0	0	n
Series 42	0	1	0	0	0	0	0	n
Series 43	2	1	1	0	0	0	-	y
Series 44	||2||	1	0	2	0	1	0	y
Series 45	1	1	0	0	0	0	2	n
Series 46	||3||	1	||1,4||	1	0	0	1	y
Series 47	||1,3||	1	0	1	0	0	||1,3||	n
Series 48	0	1	0	1	0	1	0	n
Series 49	2	1	0	1	0	0	-	n
Series 50	0	1	0	0	0	0	-	y
Series 51	0	1	0	1	0	1	0	n
Series 52	||2||	1	0	1	0	1	-	y
Series 53	1	1	0	0	0	0	-	y
Series 54	||2||	1	0	1	0	0	0	n
Series 55	1	1	0	1	0	1	0	y
Series 56	2	1	0	||2||	0	0	-	y
Series 57	1	1	0	0	0	0	-	n
Series 58	0	1	1	2	0	1	-	y
Series 59	0	1	2	1	0	0	||2||	y
Unprocessed food	0	1	0	0	0	0	0	y
Processed food	2	1	0	1	0	0	0	n
Non-energy industrial goods	||1,2,5||	1	1	1	0	1	0	y
Energy	||2||	1	0	1	0	1	-	n
Services	||3||	1	0	0	0	1	0	y
Total	1	1	0	1	0	1	0	y

Note: p - autoregressive polynomial order; d - integration order; q - moving average polynomial order; P - seasonal autoregressive polynomial order; D - seasonal integration order; Q - seasonal moving average polynomial order; R - order of the polynomial associated with the exogenous variable; ||L|| = lag of order L; y = yes; n = no.

Table 5 - Root Mean Squared Forecast Errors

	RW			SARIMA			SARIMAX					
Forecast horizon	RW	RW_5	RW_59	SARIMA	SARIMA_5	SARIMA_59	SARIMAX_F_DIR	SARIMAX_F_DIR_5	SARIMAX_F_DIR_59	SARIMAX_F_INDIR	SARIMAX_F_INDIR_5	SARIMAX_F_INDIR_59
1	0.4367	0.4529	0.4394	0.2302	0.2091	0.2069	0.2445	0.2265	0.2010	0.2374	0.2122	0.2132
2	0.7799	0.8043	0.7928	0.3529	0.3052	0.2538	0.3469	0.2774	0.2343	0.3638	0.3063	0.2554
3	1.0124	1.0432	0.9451	0.4484	0.4028	0.3220	0.4771	0.3866	0.3351	0.4790	0.4100	0.3471
4	1.1375	1.1747	1.0768	0.4922	0.4870	0.4076	0.5443	0.4816	0.4361	0.5628	0.5058	0.4513
5	1.2406	1.2858	1.2245	0.5486	0.5877	0.5420	0.6537	0.5948	0.5766	0.6551	0.6096	0.6021
6	1.4207	1.4807	1.5551	0.5713	0.6921	0.7287	0.7209	0.6805	0.8218	0.7380	0.7101	0.8114
7	1.7026	1.7740	1.8802	0.6023	0.8018	0.8561	0.8563	0.8038	0.9459	0.8515	0.8284	0.9487
8	2.0337	2.1160	2.2245	0.6624	0.9384	0.9688	1.0192	0.9817	1.1203	0.9932	0.9728	1.1154
9	2.3292	2.4281	2.4949	0.7238	1.0795	1.1264	1.1837	1.1449	1.3320	1.1280	1.1162	1.2862
10	2.5538	2.6664	2.6993	0.7612	1.2196	1.2861	1.3156	1.2836	1.4909	1.2456	1.2539	1.4554
11	2.7513	2.8718	2.8585	0.7981	1.3652	1.4356	1.4685	1.4572	1.6453	1.3751	1.4031	1.6159
12	2.9899	3.1305	3.0962	0.8296	1.5123	1.5444	1.6261	1.6259	1.8104	1.5149	1.5601	1.7497

Table 6 - RW (5 components)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.4367	0.4529	-3.5717	2.5334	0.9944
2	0.7799	0.8043	-3.0414	3.4220	0.9997
3	1.0124	1.0432	-2.9467	4.4947	1.0000
4	1.1375	1.1747	-3.1658	5.8221	1.0000
5	1.2406	1.2858	-3.5171	7.8421	1.0000
6	1.4207	1.4807	-4.0528	11.4843	1.0000
7	1.7026	1.7740	-4.0228	12.4105	1.0000
8	2.0337	2.1160	-3.8897	11.4008	1.0000
9	2.3292	2.4281	-4.0741	10.6110	1.0000
10	2.5538	2.6664	-4.2256	10.9568	1.0000
11	2.7513	2.8718	-4.1954	15.7616	1.0000
12	2.9899	3.1305	-4.4913	17.9874	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 7 - RW (59 subcomponents)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.4367	0.4394	-0.6008	0.1376	0.5547
2	0.7799	0.7928	-1.6359	0.3727	0.6453
3	1.0124	0.9451	7.1226	-1.1376	0.1276
4	1.1375	1.0768	5.6404	-0.9149	0.1801
5	1.2406	1.2245	1.3091	-0.3217	0.3739
6	1.4207	1.5551	-8.6422	3.3071	0.9995
7	1.7026	1.8802	-9.4427	3.5757	0.9998
8	2.0337	2.2245	-8.5798	4.2542	1.0000
9	2.3292	2.4949	-6.6433	4.4644	1.0000
10	2.5538	2.6993	-5.3930	4.1590	1.0000
11	2.7513	2.8585	-3.7512	5.6738	1.0000
12	2.9899	3.0962	-3.4325	8.5415	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 8 - SARIMA (5 components)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2302	0.2091	10.0821	-0.9230	0.1780
2	0.3529	0.3052	15.6485	-1.1527	0.1245
3	0.4484	0.4028	11.3145	-0.8517	0.1972
4	0.4922	0.4870	1.0759	-0.0846	0.4663
5	0.5486	0.5877	-6.6582	0.5478	0.7081
6	0.5713	0.6921	-17.4544	1.5048	0.9338
7	0.6023	0.8018	-24.8858	2.1911	0.9858
8	0.6624	0.9384	-29.4140	2.8891	0.9981
9	0.7238	1.0795	-32.9488	3.6063	0.9998
10	0.7612	1.2196	-37.5844	4.2957	1.0000
11	0.7981	1.3652	-41.5378	4.7000	1.0000
12	0.8296	1.5123	-45.1417	5.1969	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 9 - SARIMA (59 subcomponents)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2302	0.2069	11.2704	-0.9827	0.1629
2	0.3529	0.2538	39.0804	-2.2839	0.0106
3	0.4484	0.3220	39.2664	-2.3028	0.0106
4	0.4922	0.4076	20.7567	-1.4880	0.0684
5	0.5486	0.5420	1.2205	0.0997	0.4603
6	0.5713	0.7287	-21.5954	2.3055	0.9894
7	0.6023	0.8561	-29.6480	3.3554	0.9996
8	0.6624	0.9688	-31.6278	3.7077	0.9999
9	0.7238	1.1264	-35.7420	4.7780	1.0000
10	0.7612	1.2861	-40.8135	5.4452	1.0000
11	0.7981	1.4356	-44.4058	5.2139	1.0000
12	0.8296	1.5444	-46.2822	5.4878	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 10 - SARIMAX_F_DIR (5 components)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2445	0.2265	7.9376	-0.7057	0.2402
2	0.3469	0.2774	25.0442	-2.4565	0.0070
3	0.4771	0.3866	23.3919	-2.6117	0.0045
4	0.5443	0.4816	13.0266	-1.8160	0.0347
5	0.6537	0.5948	9.9044	-1.8098	0.0352
6	0.7209	0.6805	5.9381	-1.1333	0.1286
7	0.8563	0.8038	6.5300	-1.3091	0.0953
8	1.0192	0.9817	3.8188	-0.7467	0.2276
9	1.1837	1.1449	3.3879	-0.7379	0.2303
10	1.3156	1.2836	2.4936	-0.5862	0.2789
11	1.4685	1.4572	0.7740	-0.2032	0.4195
12	1.6261	1.6259	0.0114	-0.0033	0.4987

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 11 - SARIMAX_F_INDIR (5 components)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2374	0.2122	11.8704	-1.5650	0.0588
2	0.3638	0.3063	18.8006	-2.2001	0.0139
3	0.4790	0.4100	16.8248	-2.2672	0.0117
4	0.5628	0.5058	11.2577	-1.9720	0.0243
5	0.6551	0.6096	7.4609	-1.5999	0.0548
6	0.7380	0.7101	3.9296	-0.9487	0.1714
7	0.8515	0.8284	2.7911	-0.6568	0.2557
8	0.9932	0.9728	2.0990	-0.4760	0.3170
9	1.1280	1.1162	1.0618	-0.2494	0.4015
10	1.2456	1.2539	-0.6687	0.1668	0.5662
11	1.3751	1.4031	-1.9964	0.5473	0.7079
12	1.5149	1.5601	-2.9018	0.9088	0.8183

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 12 - SARIMAX_F_DIR (59 subcomponents)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2445	0.2010	21.6460	-1.6919	0.0453
2	0.3469	0.2343	48.0774	-2.9056	0.0018
3	0.4771	0.3351	42.3611	-2.6911	0.0036
4	0.5443	0.4361	24.8045	-1.8813	0.0300
5	0.6537	0.5766	13.3687	-1.3701	0.0853
6	0.7209	0.8218	-12.2797	1.6312	0.9486
7	0.8563	0.9459	-9.4786	1.4511	0.9266
8	1.0192	1.1203	-9.0293	1.6796	0.9535
9	1.1837	1.3320	-11.1354	2.2988	0.9892
10	1.3156	1.4909	-11.7569	3.2169	0.9994
11	1.4685	1.6453	-10.7494	4.1454	1.0000
12	1.6261	1.8104	-10.1795	5.0819	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Table 13 - SARIMAX_F_INDIR (59 subcomponents)

Forecast horizon	RMSFE dir	RMSFE indir	((RMSFE dir / RMSFE indir) - 1)*100	Diebold-Mariano	p-value[1]
1	0.2374	0.2132	11.3907	-0.8588	0.1952
2	0.3638	0.2554	42.4428	-2.6914	0.0036
3	0.4790	0.3471	37.9882	-2.5015	0.0062
4	0.5628	0.4513	24.7048	-1.8866	0.0296
5	0.6551	0.6021	8.8109	-1.0045	0.1576
6	0.7380	0.8114	-9.0478	1.3729	0.9151
7	0.8515	0.9487	-10.2474	1.6832	0.9538
8	0.9932	1.1154	-10.9554	2.1585	0.9846
9	1.1280	1.2862	-12.2968	2.8784	0.9980
10	1.2456	1.4554	-14.4156	4.4175	1.0000
11	1.3751	1.6159	-14.9039	6.4674	1.0000
12	1.5149	1.7497	-13.4237	7.2515	1.0000

[1] H_0: Direct forecast = Indirect forecast H_1: Indirect forecast better than Direct forecast

Figure 1 - Variance of price series explained by each principal component

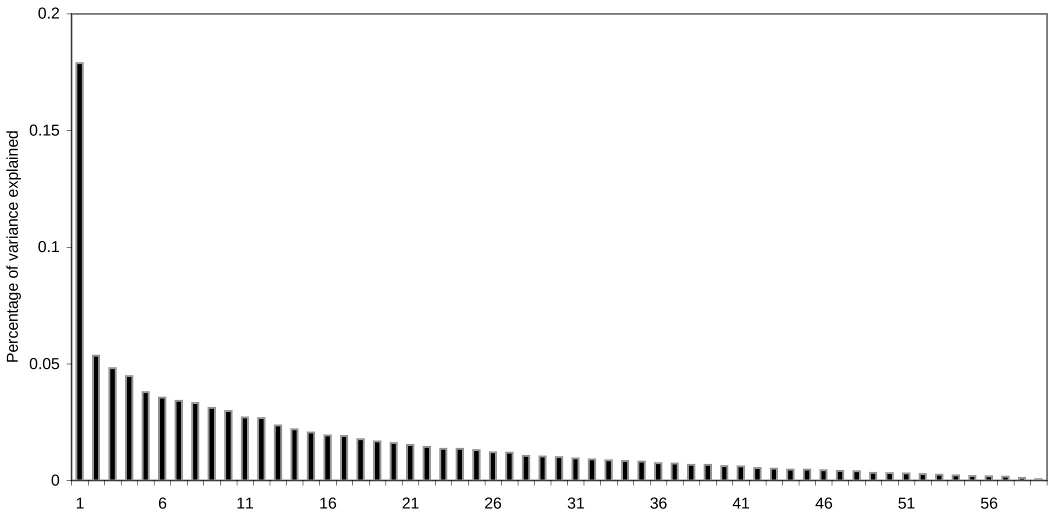

Figure 2 - Direct forecasting

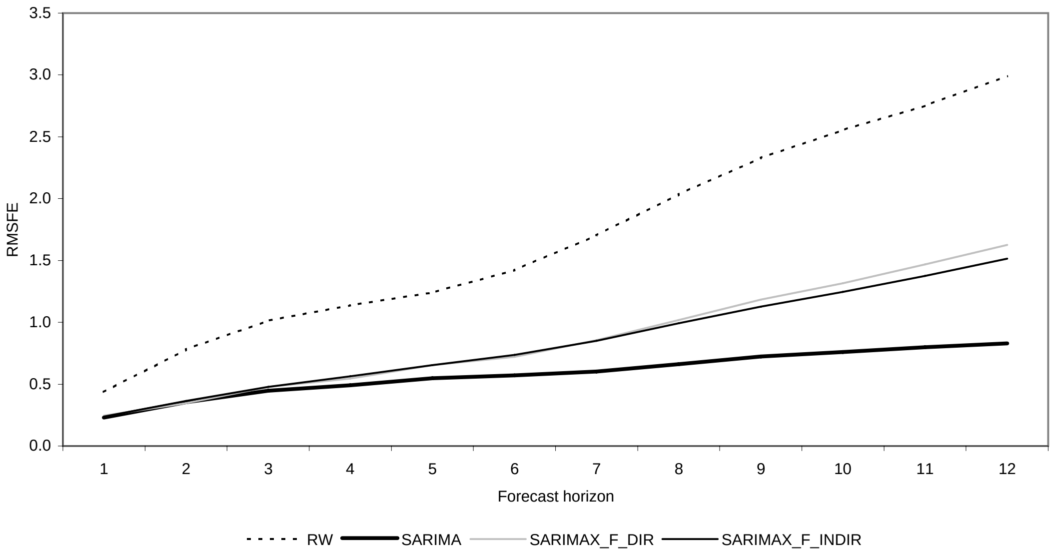

Figure 3 - Indirect forecasting (5 components)

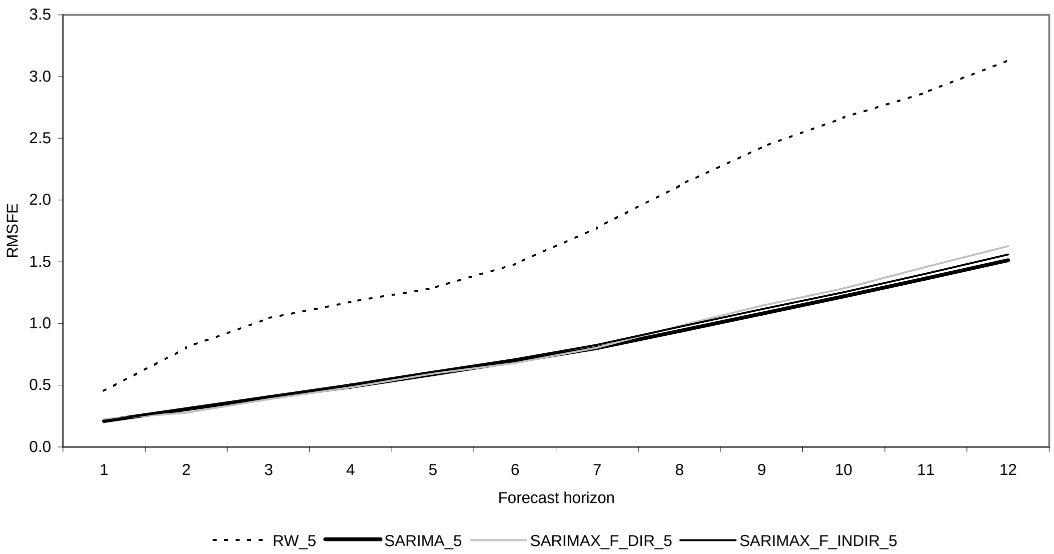

Figure 4 - Indirect forecasting (59 subcomponents)

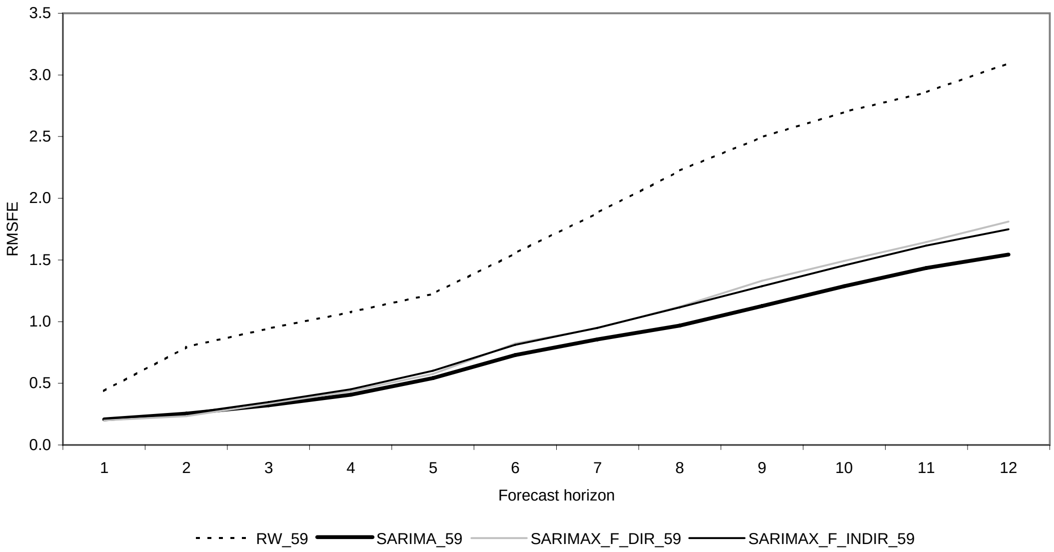

Figure 5 - Random walk

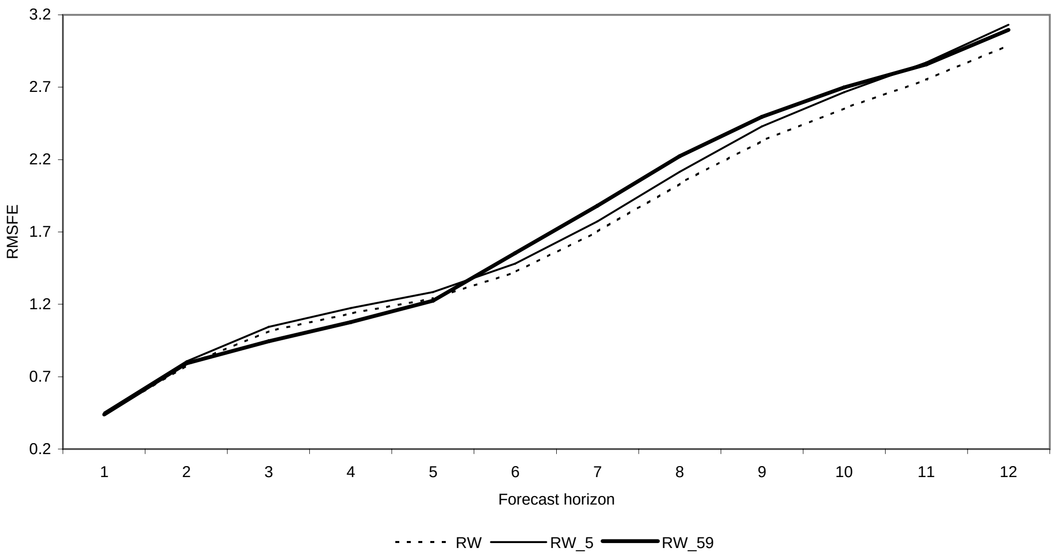

Figure 6 - SARIMA

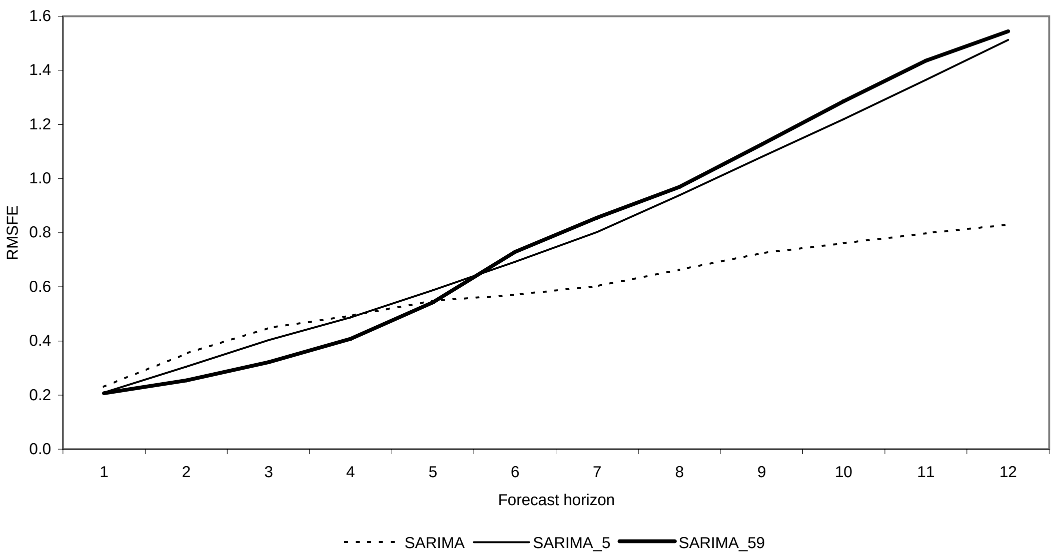

Figure 7 - SARIMAX_F_DIR

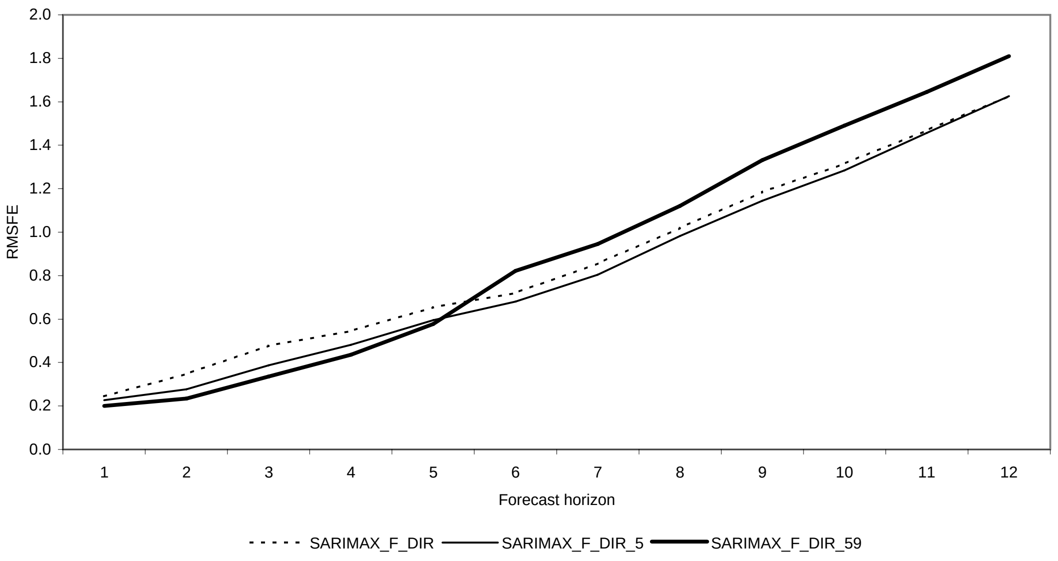

Figure 8 - SARIMAX_F_INDIR