BANKA SLOVENIJE

Prikazi in analize XIV/1 (maj 2007), Ljubljana

FORECASTING WITH ARMA MODELS
The case of Slovenian inflation

Klara Stoviček*

Abstract

The main objective of this paper is to evaluate how useful standard in-sample model selection criteria are for selecting the best model for out-of-sample forecasting of Slovenian inflation. To answer this question, a complete set of ARMA models are compared with respect to their out-of-sample forecast performance. The results are computed for various methods of seasonal adjustment and lengths of the forecast horizon. For the models with the preferred forecast ability both the in-sample mis-specification test as well as the ability to deliver the optimal forecasts are examined. All the issues of interest have been studied for monthly as well as quarterly periodicity. It has been found that in terms of forecast ability ARMA models outperform AR models, when allowing for the same degrees of freedom. Also, the models with separate specification of a seasonal component do better than models where seasonal terms are modeled jointly with other components of the time series. Eventually, the models with a trend displaying the structural break in 1999 outperform other models. Interestingly enough, in the context of the sample examined, the standard in-sample model selection criteria provide rather poor guidance in identifying the best model for out-of-sample forecasting.

JEL codes: C19, C22, E31
Keywords: out-of-sample forecasting, inflation, ARMA models, in-sample model selection criteria

Povzetek

Glavni cilj te raziskave je ugotoviti na primeru inflacije v Sloveniji, če so kriteriji za izbor modelov znotraj ocenjevalnega obdobja tudi primerni za izbor modelov za napovedovanje. V ta namen ta raziskava primerja celotno paleto ARMA modelov glede na njihovo napovedno moč, ki je odvisna tudi od metode desezoniranja in dolžine napovednega obdobja. Ugotovljeni modeli z največjo napovedno močjo so testirani za prisotnost napačne specifikacije znotraj ocenjevalnega obdobja, poleg tega pa so ovrednotene tudi njihove napovedi. Vsa ta vprašanja so analizirana na podlagi napovedovanja medmesečne in mekvartalne inflacije. Ugotovitve te raziskave so, da imajo ARMA modeli večjo napovedno moč kot AR modeli ob prilagoditvi za stopinje prostosti. Poleg tega imajo modeli z ločeno specifikacijo sezone večjo napovedno moč kot modeli, kjer sezona vstopa v regresijsko enačbo skupaj z ostalimi členi časovnih vrst. Pomemben rezultat je tudi, da imajo modeli s trendom, ki upoštevajo strukturni prelom v letu 1999, večjo napovedno moč kot modeli brez trenda. V okviru napovedovanja slovenske inflacije imajo standardni kriteriji za izbor modelov znotraj ocenjevalnega obdobja zgolj omejeno moč pri izboru modelov za napovedovanje.

JEL klasifikacija: C19, C22, E31
Ključne besede: napovedovanje izven vzorca, inflacija, ARMA modeli, kriteriji za izbor modelov znotraj vzorca

* Analysis and Research Department, Bank of Slovenia. E-mail: klara.stovicek@bsi.si.

The author thanks Damjan Kozamernik, Dejan Krušec and Matej Rojs for their discussions, the internal presentation and helpful suggestions. All remaining errors are mine. The views expressed in this article are not necessarily the views of the Bank of Slovenia.

1. INTRODUCTION

Forecasting inflation has a long tradition at the Bank of Slovenia, since it constitutes a one of necessary elements to conduct independent monetary policy. In the process of becoming a full-fledged member of the euro area, the Bank of Slovenia since autumn 2006 actively prepares the inflation forecasts four times per year in the ECB's Narrow Inflation Projection Exercise (NIPE). In forecasting, the Bank uses various types of models, in practice duly amended with expert knowledge, both those relying on structural relationships among macroeconomic variables, and atheoretical empirical properties. It is of interest to evaluate the forecasting performance of different models in particular since the theory and empirics in short-term forecasting have been undergoing substantial evolution in recent years.

Empirical work provides strong evidence that in short-term forecasting the atheoretical "non-causal" models outperform theoretical "causal" models. Clements and Hendry (1999) indicate that the refutation of the claim that causal models should outperform non-causal models is an important step towards understanding the actual behavior of economic forecasts and the value added therein. The causal or structural approach builds on the theory and in projecting inflation considers variables such as the output gap, world prices and the exchange rate. The non-causal or atheoretical approach is purely statistical and is based on the past behavior of the variable. An example of this approach is the time series approach. In forecasting competitions the latter approach has been shown to be more successful than the structural approach, in particular due to its robustness to deterministic shifts and intermittent shocks. Since the recent history of disinflation in Slovenia has been coupled with several structural breaks and level shifts, I forecast period-by-period CPI only with univariate time series models, keeping in mind the future extension to the time series multivariate models.

Two main theories have been shaping the thinking in economic forecasting. The traditional optimality theory on economic forecasting has grounds in two key assumptions, that the model is a good representation of the economy and that the structure of the economy remains unchanged also in the future. Given these assumptions, several theorems can be proved. In particular, it can be shown that that the best in-sample model produces the best forecasts. However, the empirical evidence has undermined the relevance of the two assumptions of the traditional theory together with all the related theorems. This is how the new theory on forecasting emerged, relying on the assumptions that the models are simplified representations and incorrect in many ways, and that economies are subject to sudden shifts (Clements and Hendry, 1999).

The main reason for the empirical failure of the traditional theory is that data generating process (DGP) in the past generally differs from the one in the future. Therefore, there is no evidence that the best in-sample model or equivalently the model that fits the best the past DGP provides also the best description of the future. Moreover as Diebold and Kilian (2000) show, even a correctly specified model does not necessarily need to improve the forecast accuracy relative to a mis-specified model. The winners of the forecast competition tend to be those models that are in their structure most able to adapt to the occurrence of various shocks and shifts although being a poor representation of both the economic theory and the data for the in-sample DGP.

In the more realistic framework of the new theory on economic forecasting, the forecast errors occur due to numerous reasons like shifts and mis-estimations in the coefficients of deterministic and stochastic terms, mis-specifications of deterministic and stochastic terms, the mis-measurement of the data, etc. Clements and Hendry (1999, 2001) claim that the shifts in deterministic terms are the main culprit for forecast failure, in particular since they cause a systematic forecast bias. From this perspective, the best in-sample causal models that are not robust to sudden deterministic shifts can easily lose in the forecast competition against the more adaptive, although mis-specified models able to (in part) avoid the systematic forecast failure. This finding opens the important question of whether there exists any value added in testing for mis-specifications in order to choose the best model for forecasting.

In this paper I evaluate the forecasting of the period-by-period CPI using the whole set of ARMA models, considering the monthly and quarterly frequency as well as the different methods for removing seasonality from the data. The main technique I use is the out-of-sample forecasting, which enables to study the behavior of ARMA models in a huge number of (in-sample) estimations as well as (out-of-sample) forecast samples. As model selection criteria for forecasting I consider standard criteria in the literature like the coefficient of determination (R2), the adjusted coefficient of determination (R2a), the Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC). Finally the success of models for forecasting is assessed in terms of the root mean squared forecast error (RMSFE).

The first goal of this paper is to evaluate the models according to their in-sample as well as out-of-sample forecasting performance and to find whether the criteria applied for the in-sample model selection could also be used for selecting the best model for out-of-sample forecasting. In addition, this paper directly compares the model selection criteria in terms of their forecast performance in order to find the best criterion in the in-sample estimation for selecting the models for forecasting.

I find that, in general, the models with the best in-sample as well as out-of-sample forecasting performance of the Slovenian inflation are those that include a trend. This might be seen as a statistical confirmation that the Bank of Slovenia was successful in conducting a disinflation policy, by installing a disinflation trend until inflation reaches levels consistent with price stability. Interestingly, even for the models with trend, the in-sample model selection criteria on average perform poorly in selecting the best model for forecasting, as indicates a relatively low correlation between the in-sample values of model selection criteria and the out-of-sample forecasting performance of models. In the case the period-by-period CPI is seasonally adjusted with the X12 method, the AIC generates the lowest RMSFE in the models with trend for both the monthly and quarterly periodicity.

The second goal of this paper is to identify the best models for forecasting in terms of the RMSFE and examine in the context of the obtained results two important related issues raised in the literature. One issue is whether the best models for forecasting could be mis-specified in the in-sample estimation. In other words, I want to find out whether there exists value added in mis-specifications testing for choosing the best model for forecasting. The second issue is whether the forecasts of the chosen best models for forecasting are indeed optimal in the view of several criteria, as stated for instance by Diebold and Lopez (1996). Thereby, the effort has been made also on choosing the best method of seasonal adjustment.

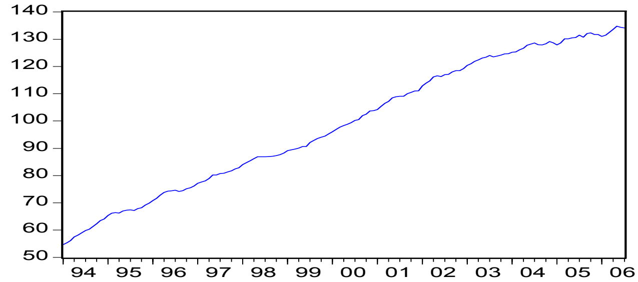
The results related to the second goal are the following. For forecasting the preferred method of removing a season is the X12 regardless of periodicity and whenever this is the case the models with trend are superior to those without trend. In addition, ARMA models tend to outperform AR models with and without trend regardless of periodicity. As expected, short-term forecasting performs on average better than medium term forecasting. Worth noting is also that the ARMA models with the lowest RMSFE tend to switch frequently between forecast horizons while this is much less the case for the AR models. Finally it has been found that the models with a superior forecast performance in general fail mis-specification check, yet deliver an unbiased and efficient out-of-sample forecast.

The organization of the paper is the following. The second section inspects the time series characteristics of the variable being forecast. The third section presents the forecasting models and the out-of sample forecast procedure. The fourth section evaluates the out-of-sample forecast performance of the in-sample model selection criteria and finally compares the criteria in terms of their forecast performance. The section fifth aims at identifying the best models for forecasting and evaluate their in-sample characteristics as well as their abilities to deliver the optimal forecast. The sixth section concludes.

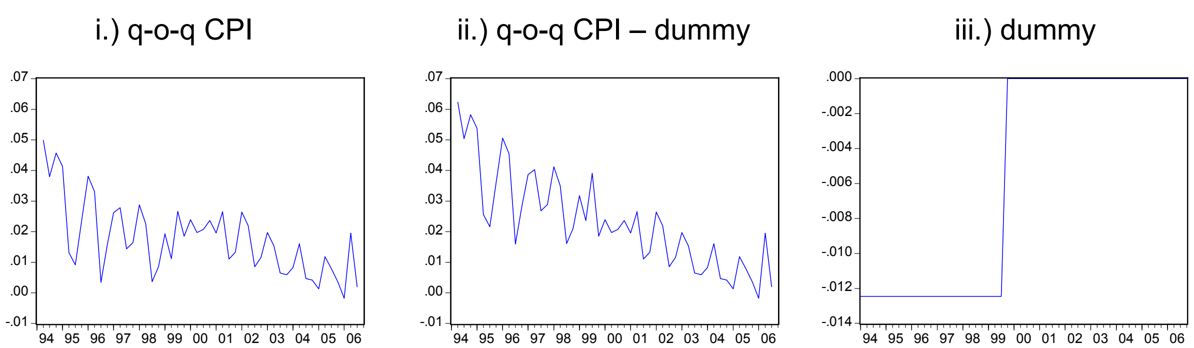
2. DATA

The data employed in this study include the consumer price index (CPI) of quarterly and monthly frequency starting in 1994(1) and ending in 2006(6) in the case of monthly data and in 2006(2) in the case of quarterly data, respectively. The CPI of a monthly frequency and the m-o-m CPI as well as the q-o-q CPI are presented in the *Figure 1, 2iv and 2i*.

Figure 1: CPI, monthly data

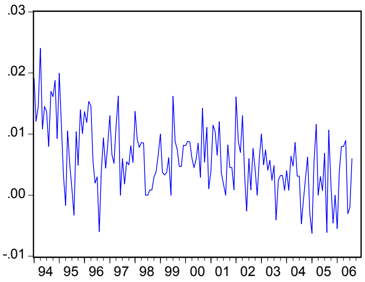


Figures 2i – 2iii: Quarterly data

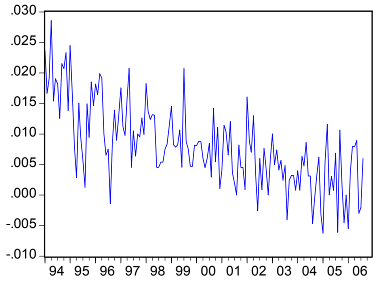


Figures 2iv – 2vi: Monthly data

iv.) m-o-m CPI



v.) m-o-m CPI – dummy



vi.) dummy

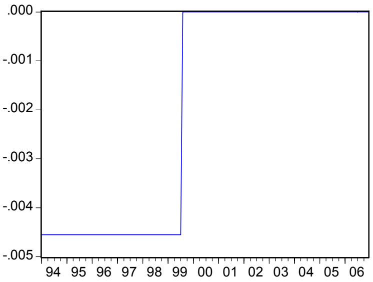
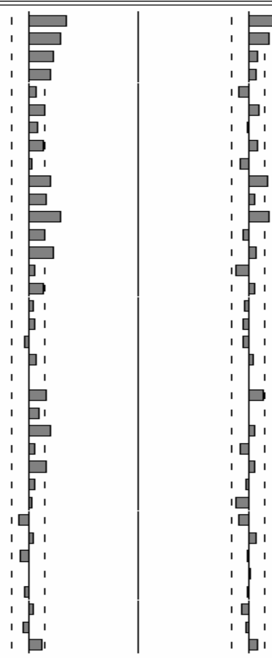


Figure 3: Correlogram of the m-o-m CPI

Sample: 1994M01 2006M07
Included observations: 151



	AC	PAC	Q-Stat	Prob
1	0.368	0.368	20.896	0.000
2	0.309	0.201	35.727	0.000
3	0.237	0.087	44.492	0.000
4	0.212	0.074	51.587	0.000
5	0.071	-0.089	52.382	0.000
6	0.155	0.102	56.201	0.000
7	0.085	-0.008	57.350	0.000
8	0.143	0.087	60.644	0.000
9	0.035	-0.068	60.845	0.000
10	0.207	0.181	67.874	0.000
11	0.166	0.064	72.397	0.000
12	0.297	0.195	87.104	0.000
13	0.154	-0.055	91.090	0.000
14	0.236	0.071	100.51	0.000
15	0.063	-0.120	101.18	0.000
16	0.141	0.054	104.57	0.000
17	0.050	-0.028	105.00	0.000
18	0.061	-0.044	105.65	0.000
19	-0.033	-0.055	105.83	0.000
20	0.081	0.044	106.98	0.000
21	0.002	0.003	106.98	0.000
22	0.177	0.133	112.60	0.000
23	0.101	0.004	114.45	0.000
24	0.210	0.054	122.47	0.000
25	0.061	-0.081	123.16	0.000
26	0.169	0.058	128.42	0.000
27	0.059	-0.022	129.08	0.000
28	0.027	-0.114	129.22	0.000
29	-0.085	-0.084	130.57	0.000
30	0.054	0.069	131.13	0.000
31	-0.070	-0.009	132.07	0.000
32	0.016	0.021	132.12	0.000
33	-0.026	-0.014	132.26	0.000
34	0.045	-0.063	132.67	0.000
35	-0.041	-0.020	133.00	0.000
36	0.127	0.083	136.24	0.000

The visual inspection of the q-o-q CPI and the m-o-m CPI as well as the correlogram of the m-o-m CPI *in Figure 3* and of the q-o-q CPI in the *Appendix 1A* display a strong seasonal pattern, given that starting with 2001 systematic seasonal movements can be detected. Therefore, the data have been seasonally adjusted using three techniques. First, the X12 seasonal adjustment method removes the cyclical seasonal movements from the series and extracts the underlying trend components of the series. As diagnostics sliding stability analysis was performed to check for the change in adjusted series over a moving sample of fixed size. Second, the seasonal adjustment with seasonal dummies assigns a dummy to each quarter or to each month, respectively, of the year, but omits the intercept term in order to avoid the dummy variable trap or the case of perfect collinearity. Thus, four seasonal dummies are considered in the case of quarterly data and 12 seasonal dummies in the case of monthly data. This method assumes that the seasonal factor, if present, is deterministic and not stochastic.[1] Third, multiplicative seasonality considers the interaction of ARMA and seasonal effects such that in regression in addition to the constant term, enter the seasonal autoregressive (SAR) and seasonal moving average (SMA) terms for certain lags. Inspecting the Figure 3 of the m-o-m CPI reveals that the autocorrelation coefficients decline and then rise to a peak at lag 12 while the partial autocorrelation coefficients display spikes at lags 1, 2 and 12. This pattern suggests SAR(12), SMA(1) and SMA (12) terms as an approximation for capturing the season. The correlogram of the q-o-q CPI is displayed in the *Appendix 1A* and suggests SAR(4), SMA(1) and SMA(4) terms. Considering SAR and SMA terms imposes a nonlinear restriction on the coefficients in the estimation.

In order to test for non-stationarity in univariate models it is worth applying ADF test, see Diebold and Kilian (2000). Thus, the Augmented Dickey Fuller (ADF) test has been carried out for the q-o-q CPI and the m-o-m CPI. Table 1 presents the results of this test. Non-stationarity is rejected for the longest recursive estimation sample (1994(1)-2006(2) for quarterly data and 1994(1)-2006(6) for monthly data) considering a non-zero mean or both, a non-zero mean as well as a linear trend, taking into account also seasonal adjustment with seasonal dummies. The same result holds for carrying out the ADF test for the q-o-q CPI and the m-o-m CPI seasonally adjusted with the X12 method. The result is not surprising since the seasonal adjustment per se cannot eliminate a time-varying mean in the case it existed. The table with the results of this test is given in the *Appendix 1B*.

Table1: ADF test of the q-o-q CPI and m-o-m CPI:

Series	Period	**None-zero mean**		**Non-zero mean + trend**		**None-zero mean + trend + seasonal dummies**	
		ADF	lags (AIC,SBC)	ADF	lags (AIC, SBC)	ADF	lags (AIC, SBC)
q-o-q CPI	1994Q1-2006Q2	-3.6751***	3(3,3)	-4.0400***	3(3,0)	-4.5017***	0(0,0)
m-o-m CPI	1994M1-2006M6	-5.5329***	1(1,0)	-5.8390***	4(4,0)	-5.0592***	2(2,0)

*** indicate a rejection of the null hypothesis of the unit root at the 1% level.

The number of lags outside parenthesis is the actually used number of lags in the test; the numbers in parenthesis are the lags suggested by AIC and SBC, respectively.
For quarterly data the number of observations is at the limit that makes sense for carrying out ADF test.

The visual inspection of the q-o-q CPI and the m-o-m CPI and testing for the structural break indicate the presence of a structural break. Testing for the structural break has been executed for the breaking time of the 3rd quarter and the 7th month of 1999. As shown in the

[1] The dummy variables technique is an appropriate method to deseasonalize a time series as long as the seasonal, the trend, the cyclical and the random components enter additively the time series rather than multiplicatively.

tables 2i and 2ii the null hypothesis of no structural break is rejected for both suggested breaking times by the F-test as well as the Likelihood Ratio test. In order to adjust for those two breaks a dummy variable as shown in the Figure 2iii for the quarterly data and in the Figure 2vi for the monthly data is subtracted from the original q-o-q CPI in the Figure 2i and from the original m-o-m CPI in the Figure 2iv, respectively. The adjusted q-o-q CPI and the m-o-m CPI obtained in this way are displayed in Figures 2ii and 2v, respectively. In the *Appendix 1C* a dummy variable for both periodicities is defined.

Table 2i: Quarterly data:

Chow Breakpoint Test: 1999Q3

F-statistic	6.850442	Probability	0.002488
Log likelihood ratio	13.03527	Probability	0.001477

Table 2ii: Monthly data

Chow Breakpoint Test: 1999M07

F-statistic	7.400275	Prob. F(2,147)	0.000867
Log likelihood ratio	14.48571	Prob. Chi-Square(2)	0.000715

3. FORECASTING MODELS AND FORECAST PROCEDURE

My analysis involves a forecast competition using the whole set of ARMA models with AR lags ranging from 1 to 4 and MA terms ranging from 0 to 4, thus nesting pure AR specifications. In addition, I consider the model specifications with and without trend, bringing the total number of models to 40.[2] Overall, I have a collection of 4 major groups of models: pure AR models with trend (ARt) and without trend, and ARMA models with trend (ARMAt) and without trend. For the estimation and forecast evaluation of the period-by-period CPI I consider 3 different methods of seasonal adjustment thus increasing the overall number of ARMA models to 120.

Using data on the q-o-q CPI and the m-o-m CPI each of the 120 ARMA models is estimated with a rolling split of the in-sample estimation and out-of-sample forecast samples.[3] Each forecast sample is evaluated for various forecast horizons, from 1 to 4 steps (quarters) ahead for the quarterly data and from 1 to 12 steps (months) ahead for the monthly data.[4] The purpose of having different forecast horizons is to check whether superior forecasting models and criteria change from the short term to the medium term forecast horizon.

[2] The models considered are ARMA(1,0), ARMA(2,0), ARMA(3,0), ARMA(4,0), ARMA(1,1), ARMA(2,1), ARMA(3,1), ARMA(4,1), ARMA(1,2), ARMA(2,2), ARMA(3,2), ARMA(4,2), ARMA(1,3), ARMA(2,3), ARMA(3,3), ARMA(4,3), ARMA(1,4), ARMA(2,4), ARMA(3,4), ARMA(4,4), ARMA(1,0,t), ARMA(2,0,t), ARMA(3,0,t), ARMA(4,0,t), ARMA(1,1,t), ARMA(2,1,t), ARMA(3,1,t), ARMA(4,1,t), ARMA(1,2,t), ARMA(2,2,t), ARMA(3,2,t), ARMA(4,2,t), ARMA(1,3,t), ARMA(2,3,t), ARMA(3,3,t), ARMA(4,3,t), ARMA(1,4,t), ARMA(2,4,t), ARMA(3,4,t), ARMA(4,4,t)

[3] I consider both periodicities due to the problem of degrees of freedom in the case of quarterly data.

[4] Forecast horizon is the length of time from the period in which the forecast is made to the period being forecast.

Regarding the data on the q-o-q CPI, the first estimation sample starts in 1994(1) and ends in 2001(4) in order to produce forecasts for the period 2002(1) to 2002(4). The next estimation sample is extended for one quarter up to 2002(1) in order to produce the forecasts for the period 2002(2) to 2003(1). This rolling procedure continues until the last 15th recursive estimation sample extending from 1994(1) up to 2005(2) to generate forecasts for 2005(3) up to the end of the data sample range 2006(2). Similarly, considering the data on the m-o-m CPI, an estimation is carried out on the basis of 43 recursive estimation samples starting from 1994(1) up to 2001(12), extending the sample by one month sequentially. The longest recursive estimation sample ends in 2005(6) in order to produce forecasts for 2005(7) up to the end of the data sample range 2006(6). Thus, the whole exercise produces for all considered ARMA specifications (120*15) estimation and forecast samples for the quarterly data and (120*43) estimation and forecast samples for the monthly data thereby considering that each forecast sample is evaluated for all respective forecast horizons.

For each estimation sample, I collect the values of the model selection criteria, such as the coefficient of determination (R2), the adjusted coefficient of determination (R2a), the Akaike information criterion (AIC) and the Schwartz Bayesian criterion (SBC). These model selection criteria weight the quality of the model's in-sample fit against the degrees of freedom the model uses for this fitting performance. Therefore they minimize the in-sample residual sum of squares, but, except for R2, impose a penalty for including additional regressors in the regression. Moreover to check for the presence of autocorrelation I store the values of the Q-statistics for all recursive estimation samples. Similarly, for each forecast sample I collect the forecast errors for each of the 1 to 4 step-ahead forecast horizons in the case of quarterly data and for 1 to 12 step-ahead forecast horizons in the case of monthly data.

In addition to the quarterly and monthly models with the AR and MA terms up to the lag 4, I consider for the monthly data the group of the pure AR models for which I increase the number of AR lags from 4 to 12. As a method of removing a season from the m-o-m CPI I consider only the X12 method and the method of seasonal dummies. I cannot apply the multiplicative method due to a (already) large number of AR terms. The reason for this selection of the group of models is to check whether the results from AR models with 4 lags change with increased number of lags.

Since I estimate all possible nested ARMA models up to the lag 4, many of the models may suffer from some kind of model specification error, the most important one probably being the omission of the relevant regressor (AR or MA term) as well as the inclusion of an unnecessary regressor in the model. In the first case or the case of underfitting, the models suffer from the biased, yet in general more efficient, parameter estimates. In the second case or the case of overfitting, the models retain the unbiased parameter estimates, though at the loss in the efficiency of estimators that may be exacerbated by the problem of multicollinearity among regressors.

Economic forecasts typically differ from the actual outcome reflecting forecast uncertainty. The literature on forecasting offers numerous measures for evaluating the forecast uncertainty with the root mean square forecast error (RMSFE) being the most common measure. RMSFE combines inherently in its structure the bias and the variance of the forecast and it simplifies to the standard deviation of the forecast error in the case of unbiased forecast. The problem of the RMSFE as a measure of uncertainty is that the distributions of forecast errors might have the same variance but different densities (Hendry, Ericsson 2001). This problem limits the comparability across the models in terms of this measure.

The RMSFE evaluation is based on recursive forecasts and involves an average of the respective horizon squared forecast errors over all recursive forecast samples. Specifically, the RMSFE for the 1-step-ahead forecast (RMSFE1) of the q-o-q CPI for the certain ARMA specification is calculated as a squared root of an average of squared forecast errors over all 15-recursive forecast samples. Denote the forecast sample as t= T+1, T+2, … T+15, y_t as the actual value of dependant variable and $\hat{y}_t$ as the 1-step-ahead forecasted value in period t, then the formula for RMSFE1 is the following:

$$RMSFE1 = \sqrt{\sum_{t=T+1}^{T+15} (\hat{y}_t - y_t)^2 / 15}$$

While comparing the RMSFE of the models used to forecast the series with a different method of seasonal adjustment it is important in the first step to establish this comparability. Specifically, if I remove the seasonal component out of the variable, I forecast it independently ($\hat{y}_{ts}$) using the particular method of seasonal pattern and include it in the calculation of the RMSFE1 in the following way:

$$RMSFE1 = \sqrt{\sum_{t=T+1}^{T+15} (\hat{y}_t + \hat{y}_{ts} - y_t)^2 / 15}$$

In this way adjusted RMSFE enables the comparability of the models used to forecast the period-by-period CPI that was seasonally adjusted using different methods.

Clements and Hendry (1999) formally establish that when the DGP is susceptible to structural breaks, introducing an intercept term in the regression could mitigate systematic forecast failure. In our case, since the period-by-period CPI faces various shocks prior to the forecast origin an intercept is introduced in each of the ARMA regressions in order to capture any fixed, systematic shift in the series between the subsequent periods (except in the case when the period-by-period CPI is seasonally adjusted with seasonal dummies). In particular, an intercept correction places the model back on track at each forecast origin and enables to set the most recent ex-post forecast error to zero.

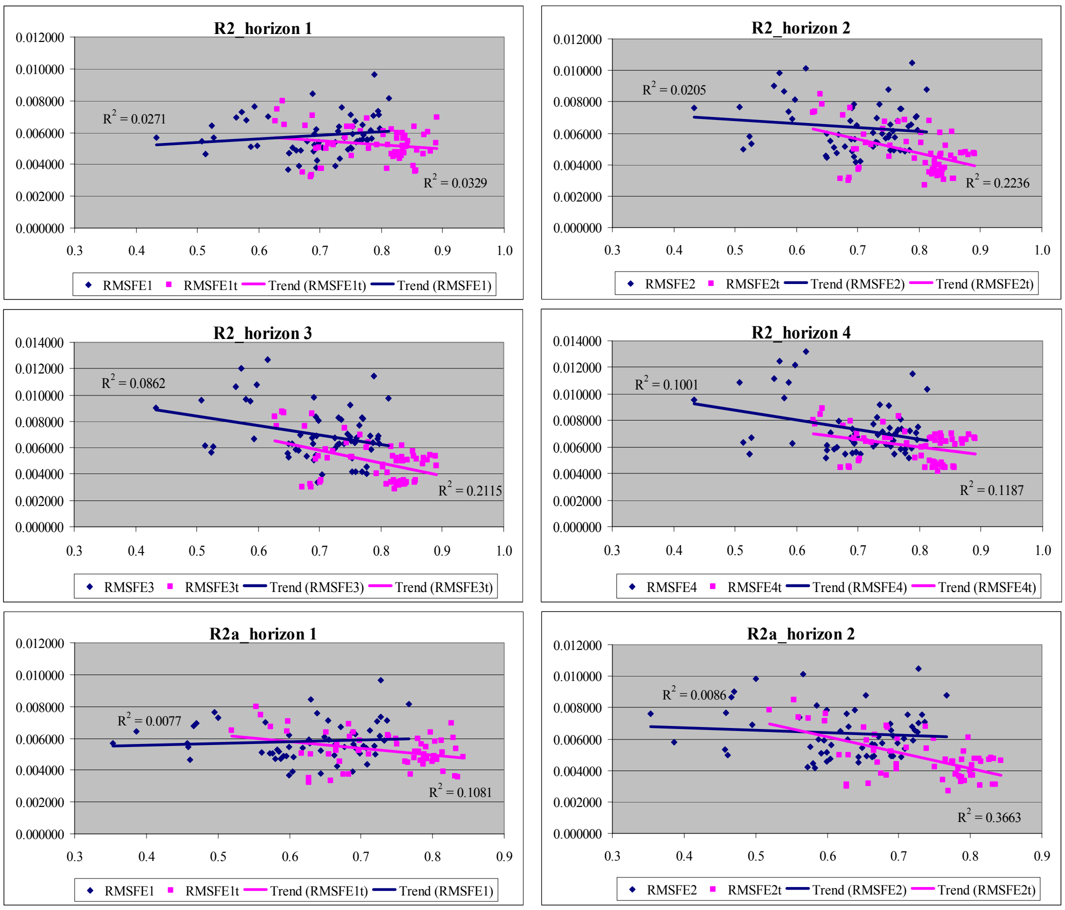
The forecast function of ARMA satisfies the difference equation that can be solved by the method of undetermined coefficients. As long as the roots of the difference equation lie within the unit circle, the forecast converges to the unconditional mean. Solving for the forecast function involves many steps. The first step is supposed to find all homogenous solutions. The second step intends to find the particular solution. The third step intends to form the general solution as the sum of the homogenous and particular solutions. The fourth step goes on to impose the initial condition and finally to rewrite the solution in the form of a t-step ahead forecast function. Although the forecasts from the ARMA specifications in the theory are unbiased, they are inaccurate since the variance of the forecast error appears as an increasing function of the length of the forecast horizon. This fact is consistent with the nature of inflation that is characterized by strong time dependence or persistence. Accordingly short-term forecasts should display superior forecast performance and I expect in the empirical exercise that 1-step-ahead forecast will outperform the forecasts of the subsequent forecast horizons in terms of RMSFE.

4. EVALUATION OF THE STANDARD IN-SAMPLE MODEL SELECTION CRITERIA

The out-of-sample performance of in-sample model selection criteria and their comparison

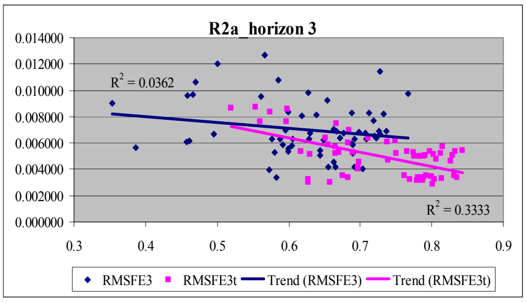
The purpose of the first empirical exercise is to evaluate the performance of in-sample model selection criteria in out-of-sample forecasting. For that I plot for each of 120 models its RMSFEs (for the 1st (RMSFE1), 2nd (RMSFE2), 3rd (RMSFE3) and 4th (RMSFE4) step-ahead forecast horizon in the case of quarterly data and for the 1st (RMSFE1), 4th (RMSFE2), 8th (RMSFE3) and 12th (RMSFE4) step-ahead forecast horizon in the case of monthly data) against the average value (over all recursive estimation samples) of the respective model estimation criteria R2, R2a, AIC and SBC. I also divide all 120 models in two groups of 60 models, with and without trend in a regression. For both groups of models I draw the linear trend fitting the observations in order to find whether there exists any correlation between the in-sample values of the model selection criteria and the out-of sample forecast performance of models in terms of RMSFE. All the results are presented in the scatterplots for both quarterly and monthly periodicities in the Figures 4i - 4xvi and Figures 5i - 5xvi, respectively.

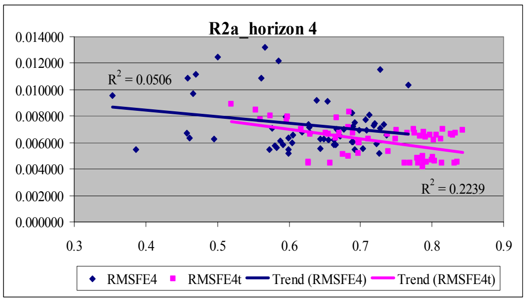
Figures 4i - 4xvi: Quarterly data[5]

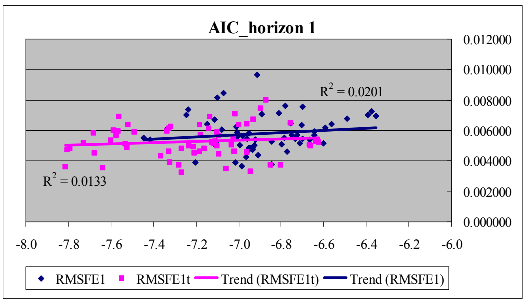


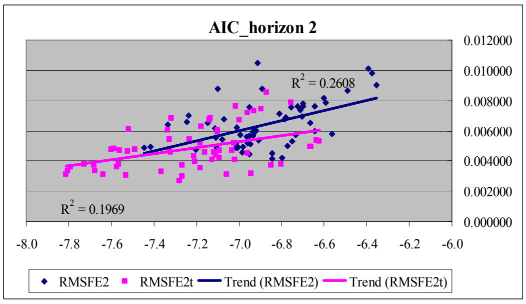
---

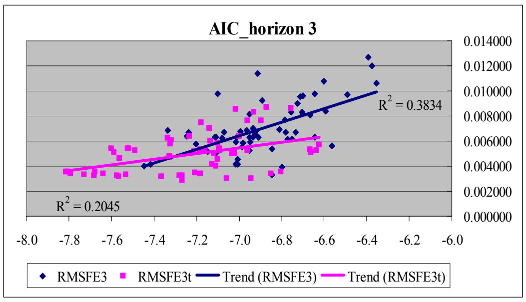
5 Each point in the scatterplot is a certain model with its average value of the respective model selection criterion over all recursive estimation samples on the x axis and the respective RMSFE on the y axis.

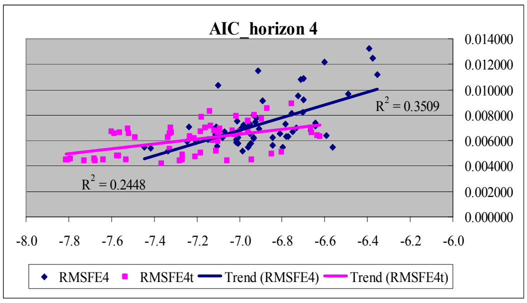

R2a_horizon 3
R² = 0.0362
R² = 0.3333
RMSFE3 RMSFE3t Trend (RMSFE3) Trend (RMSFE3t)

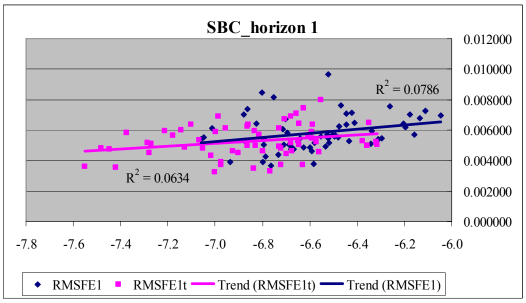

R2a_horizon 4
R² = 0.0506
R² = 0.2239
RMSFE4 RMSFE4t Trend (RMSFE4) Trend (RMSFE4t)

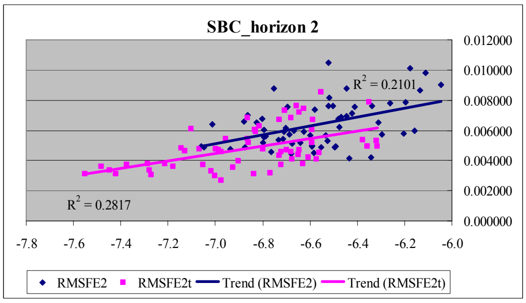

AIC_horizon 1
R² = 0.0201
R² = 0.0133
RMSFE1 RMSFE1t Trend (RMSFE1t) Trend (RMSFE1)

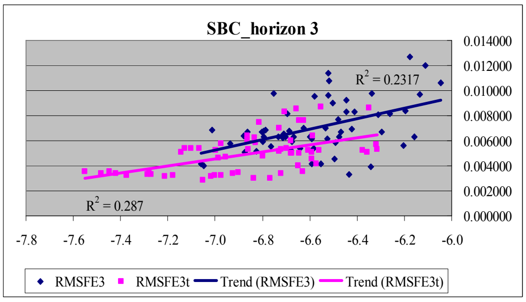

AIC_horizon 2
R² = 0.2608
R² = 0.1969
RMSFE2 RMSFE2t Trend (RMSFE2) Trend (RMSFE2t)

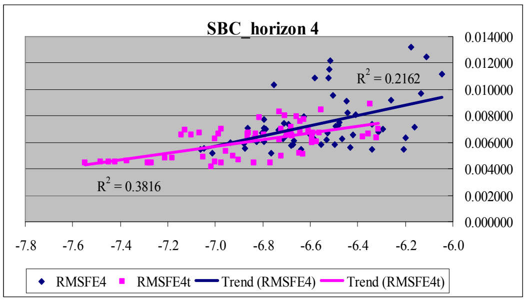

AIC_horizon 3
R² = 0.3834
R² = 0.2045
RMSFE3 RMSFE3t Trend (RMSFE3) Trend (RMSFE3t)


AIC_horizon 4
R² = 0.3509
R² = 0.2448
RMSFE4 RMSFE4t Trend (RMSFE4) Trend (RMSFE4t)

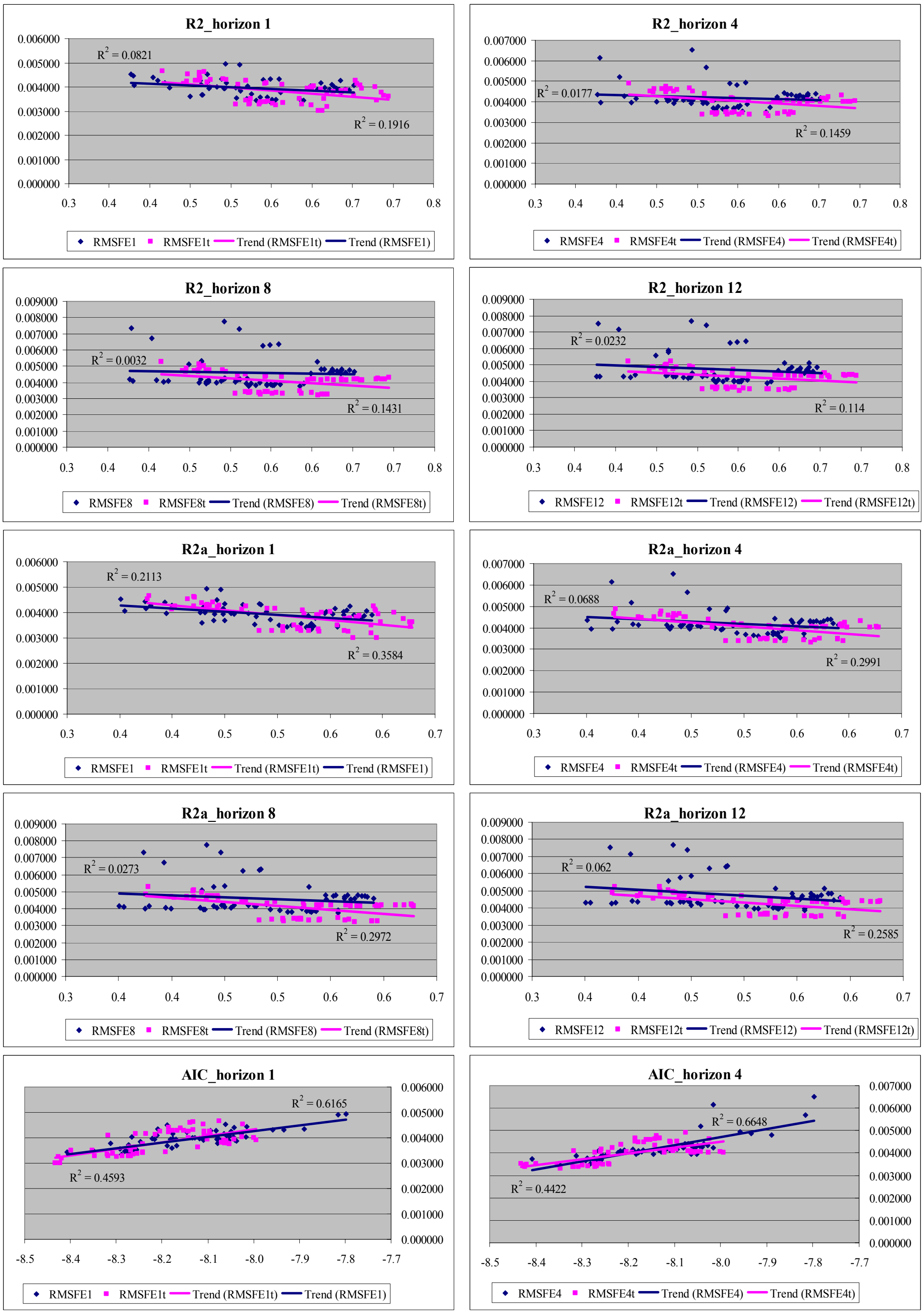

SBC_horizon 1
R² = 0.0786
R² = 0.0634
RMSFE1 RMSFE1t Trend (RMSFE1t) Trend (RMSFE1)

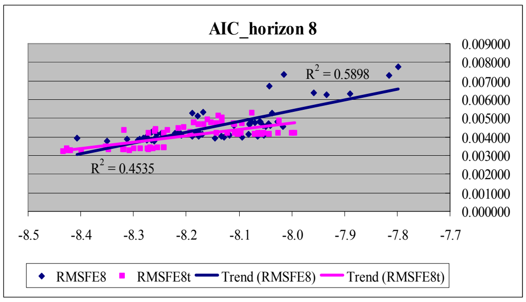

SBC_horizon 2
R² = 0.2101
R² = 0.2817
RMSFE2 RMSFE2t Trend (RMSFE2) Trend (RMSFE2t)

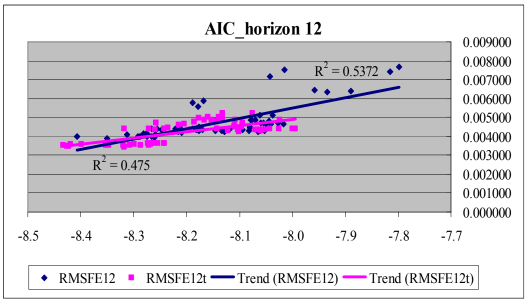

SBC_horizon 3
R² = 0.2317
R² = 0.287
RMSFE3 RMSFE3t Trend (RMSFE3) Trend (RMSFE3t)

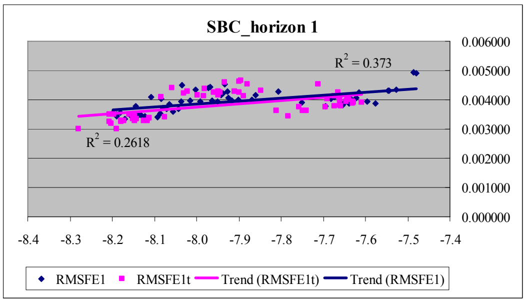

SBC_horizon 4
R² = 0.2162
R² = 0.3816
RMSFE4 RMSFE4t Trend (RMSFE4) Trend (RMSFE4t)

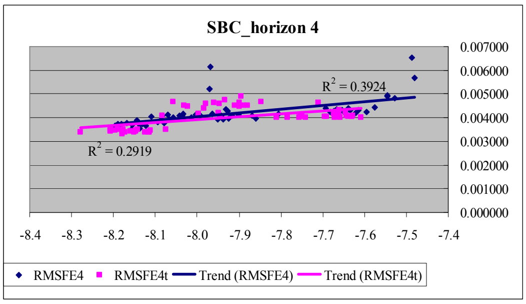
Figures 5i - 5xvi: monthly data

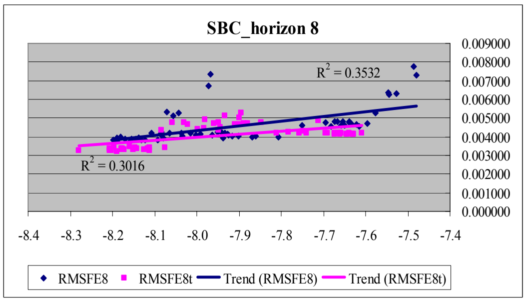


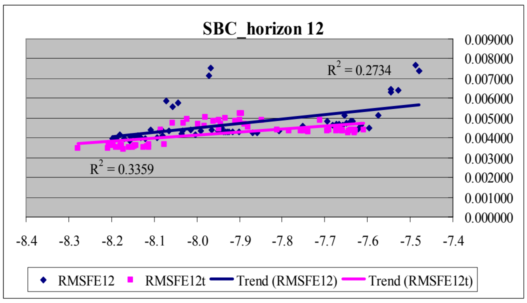












Inspecting the scatterplots offers some immediate findings for both the monthly and quarterly periodicity. First, models with trend included in a regression have in most of the cases on average the lower RMSFE and better in-sample values of model selection criteria, R2, R2a, AIC and SBC, than models without a trend. This indicates that model specifications with trend on average do better than the model specifications without trend for the in-sample estimation as well as for the out-of-sample forecast. Second, in most of the cases there exists an expected correlation between the in-sample values of the model selection criteria and the out-of-sample forecast performance of models, indicating that the models with the better in-sample values of models selection criteria on average tend to have also the lower RMSFE. Third, for the models without a trend the correlation between the in-sample values of R2 and R2a and RMSFE seems to be very loose and considerably better for AIC and SBC for most of forecast horizons. As a result, for the purpose of forecasting, R2 and R2a should not be used as selection criteria for the models without a trend. Better alternatives are the AIC and the SBC, in particular the AIC criterion for the models with the monthly data. Fourth, models with trend display somehow higher correlation between the in-sample values of model selection criteria, R2 and R2a, and RMSFE. In particular, for the quarterly periodicity R2a and SBC seem to be the most appropriate model selection criteria for forecasting while for the monthly periodicity AIC beats all the other criteria. However, even for our (superior) model specifications with trend, the proportion of the variation in the RMSFE that is explained with the variation in the in-sample values of the AIC (the superior model criterion) never exceeds 50%. This result is most likely due to a data generating process (DGP) that characterizes the Slovenian disinflation history since it differs in the estimation and forecast period. This result on the one hand supports the finding in the literature that the best in-sample model does not necessarily produce the best forecast. On the other hand, this result indicates that in-sample model selection criterion might not be the most appropriate for choosing the models for forecasting in short samples or in samples subject to shifts in the DGP. Perhaps some other measures would do a better job in the model selection for forecasting.

I continue with the second exercise to find which model selection criterion has the best forecast performance as measured by RMSFE.[6] The exercise proceeds as follows. In the first step, for each recursive estimation sample the models with the best values of in-sample model selection criteria (within each group of the models) are chosen together with their forecast errors for all forecast horizons. Interesting to note at first, for the group of AR models the various model selection criteria tend to choose the same best in-sample models while for the group of ARMA models the model selection criteria tend to choose different best in-sample models. In the second step, for each model selection criterion the RMSFE is calculated for all forecast horizons using those forecast errors of the best in-sample models over all recursive samples. Recall that there are 15 and 43 recursive forecast samples for quarterly and monthly data, respectively. In the third step, the criteria are compared according to their RMSFE for each forecast horizon where a lower RMSFE is preferred. As a benchmark for a comparison I take the RMSFE of the SBC criterion. The graphs in the *Appendix 2A* show the results within each group of the models used to forecast the q-o-q CPI and the m-o-m CPI seasonally adjusted with X12 method, seasonal dummies and multiplicative method.

The general finding is that the method of seasonal adjustment importantly affects the selection of the criterion with the lowest RMSFE. Interestingly however, within each group of models for quarterly data the criteria with the lowest RMSFE tend to remain the same for all forecast horizons regardless of the method of seasonal adjustment, while for monthly data the criteria with the lowest RMSFE tend to change somewhat more often.

Using data on the q-o-q CPI, seasonally adjusted with the X12 method, the AIC appears to be the one with the lowest RMSFE for all the models with and without trend over most of the forecast horizons. For the m-o-m CPI, seasonally adjusted with the X12 method, the AIC remains the favorite model selection criterion for forecasting in the group of the AR models with and without trend for all forecast horizons. However for the remaining groups of models, ARMA models with and without trend, AIC and SBC have the lowest RMSFE only for the 4th and 12th step-ahead forecast while R2 has the lowest RMSFE for the 1st and 8th step-ahead forecast. In the monthly specification with 12 AR lags the AIC should be used as the model selection criterion for forecasting in both groups of models, with and without trend for all forecast horizons.

For the q-o-q CPI and the m-o-m CPI, seasonally adjusted with seasonal dummies, the AIC and SBC tend to be the favorite criteria for selecting the models for forecasting in the groups of AR, ARt and ARMAt models. R2 is also the criterion with the lowest RMSFE for quarterly ARt while R2a tends to select the models with the best forecast performance in the quarterly ARMA group of models. No conclusion about the best model selection criterion can be made for both monthly groups of AR models with and without trend that consider 12 AR lags.

Regarding the period-by-period CPI, seasonally adjusted with the multiplicative method, for purposes of selecting the models for forecasting among ARMA and ARMAt groups of

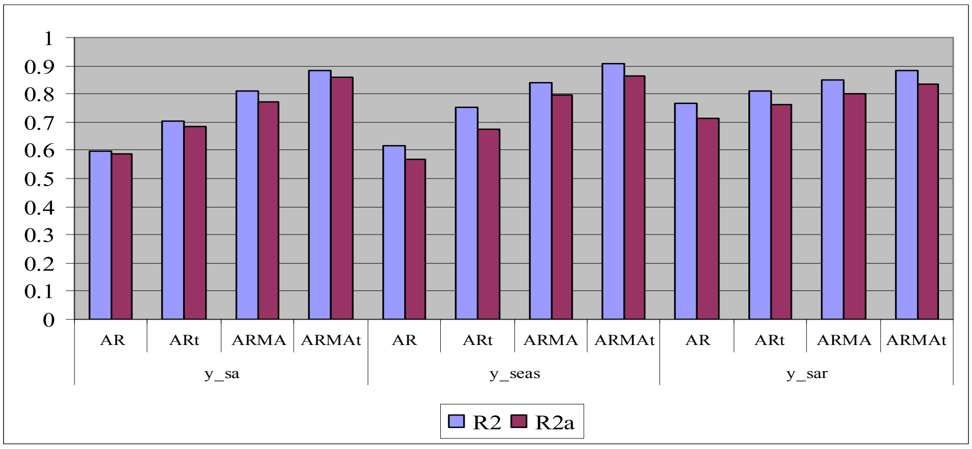
[6] In this paper I am searching only for the criterion with the lowest RMSFE. This criterion is definitely among the best criteria, although by now I have not done any statistical test (as proposed by Diebold and Mariano (1995), Harvey et al (1997)) to check whether this criterion is also statistically better from the other criteria with the higher RMSFE.

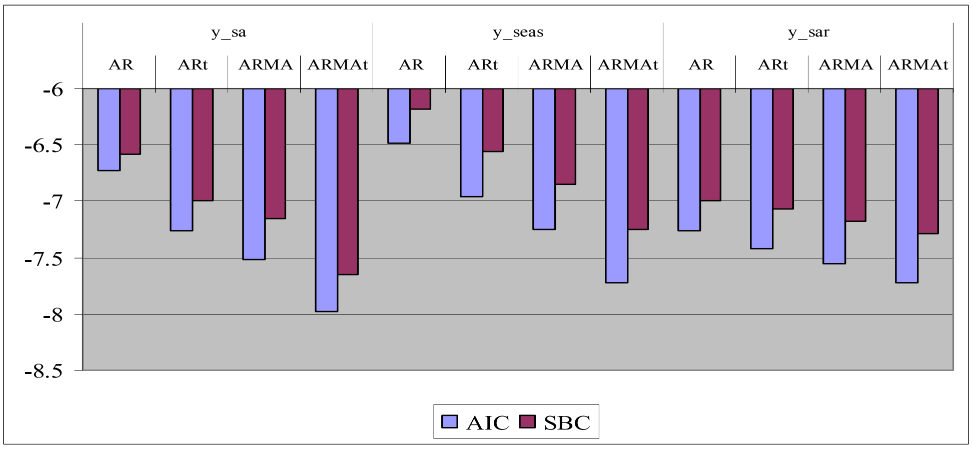
models, AIC and SBC have the lowest RMSFE in monthly data while R2 tends to be preferred for quarterly data. For quarterly data, SBC is the preferred criterion for selecting the models for forecasting in the AR group of models while for the ARt group of models the decision about the best criterion is inconclusive for both periodicities.

The in-sample behavior of model selection criteria

The third exercise is intended to emphasize the in-sample behavior of model selection criteria across the 4 groups of models (AR, ARt, ARMA, ARMAt) regressed on the q-o-q CPI and the m-o-m CPI, seasonally adjusted with X12 method, seasonal dummies and multiplicative method. Specifically, I compare across the 4 groups of models the averages of the in-sample values of the model selection criteria where those in-sample values are the best values (within each group of models) from each recursive estimation sample. Recall that there are 15 and 43 recursive forecast samples for quarterly and monthly data, respectively.

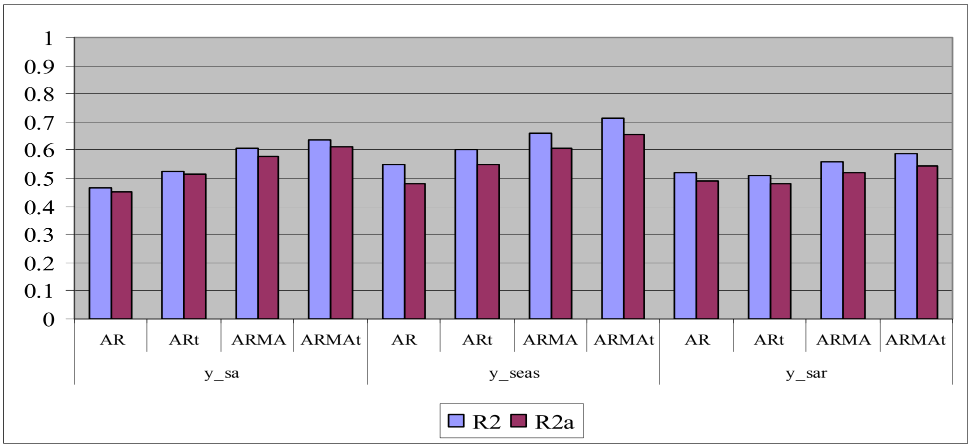
Figures 6i – 6ii: Quarterly data

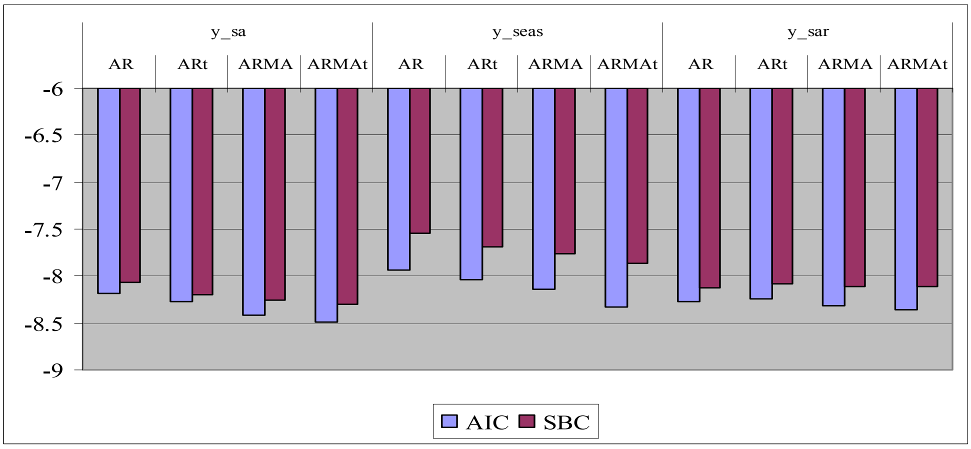




y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression
y_sar denotes the period-by-period CPI seasonally adjusted with the multiplicative method

Figures 6iii – 6iv: Monthly data





y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression
y_sar denotes the period-by-period CPI seasonally adjusted with the multiplicative method

Two immediate findings could be observed from the figures 6i – 6iv. First, in most of the cases the best values of all in-sample model selection criteria on average improve with adding to AR models a trend and MA terms for monthly and quarterly periodicity, regardless of the method of removing seasons. Concretely, in most of the cases models with trend included have on average superior best values of in-sample model selection criteria than models without a trend for both periodicities and regardless of the method of removing seasons. This result holds also for the pure AR specification with the 12 lags as it could be seen from the figures in the *Appendix 2B*. Similarly, in most of the cases ARMA models with and without trend have superior best values of in-sample model selection criteria than AR models with and without trend for monthly and quarterly data regardless of a method of removing a season.

Second, the comparison of best values of in-sample criteria across the periodicities show that models regressed on the q-o-q CPI have on average higher R2 and R2a than models regressed on the q-o-q CPI while the opposite is the case on average for AIC and SBC.

5. EVALUATION OF MODELS FOR FORECASTING

The main goal of this section is to identify the models with the lowest RMSFE and evaluate their in-sample characteristics as well as their abilities to deliver the optimal forecast.[7] I also compare different model specifications in order to obtain the best method of seasonal adjustment.

The identification of the models with the superior forecast performance

In the first exercise I want to identify the models that tend to have the best ability to forecast the period-by-period CPI. For that I order the models within the 4 groups of models (AR, ARt, ARMA, ARMAt) according to their RMSFEs, taking into account different methods of seasonal adjustment and different forecast horizons. The tables in the *Appendix 3A* report in detail the models and corresponding lowest RMSFEs for both periodicities. *The tables 3i and 3ii* are just a short summary of this collection of the models picking among the 4 groups of models for each forecast horizon just the one with the lowest RMSFE.

Table 3i: Summary table of the models with the lowest RMSFE, quarterly data

	Model	RMSFE1	model	RMSFE2	model	RMSFE3	Model	RMSFE4
y_sa	AR(4,t)	0.003235	ARMA(2,3,t)	0.002703	ARMA(1,2,t)	0.002851	ARMA(2,4,t)	0.004162
y_seas	ARMA(2,1,t)	0.003917	ARMA(1,2,t)	0.004086	ARMA(4,4,t)	0.004582	ARMA(2,4,t)	0.006181
y_sar	AR(1)	0.004643	ARMA(4,4,t)	0.004105	ARMA(3,3)	0.003343	ARMA(3,1)	0.005183

Table 3ii: Summary table of the models with the lowest RMSFE, monthly data

	Model	RMSFE1	model	RMSFE4	model	RMSFE8	Model	RMSFE12
y_sa	ARMA(2,2,t)	0.003011	ARMA(2,3,t)	0.003306	ARMA(4,4,t)	0.003232	ARMA(3,1,t)	0.003436
y_seas	ARMA(2,2,t)	0.003434	ARMA(1,3,t)	0.004014	ARMA(3,3,t)	0.004129	ARMA(2,4,t)	0.004270
y_sar	ARMA(3,1)	0.003852	ARMA(4,1)	0.003931	ARMA(4,1)	0.003932	ARMA(2,3)	0.004181

y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression
y_sar denotes the period-by-period CPI seasonally adjusted with the multiplicative method
RMSFE1 = root mean squared forecast error for the 1st forecast horizon
RMSFE2 = root mean squared forecast error for the 2nd forecast horizon etc.

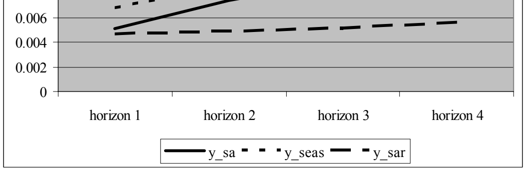
The findings are the following. First, the way of removing the season is important for the selection of the models with the lowest RMSFE. For forecasting period-by-period CPI seasonally adjusted with the X12 and with the seasonal dummies the lowest RMSFE have the models that include a trend for all forecast horizons, while the models without a trend do better for forecasting period-by-period CPI seasonally adjusted with the multiplicative method. The models with trend do better for forecasting also in the group of pure AR models with 12 AR lags, as it could be readily seen from the summary and disaggregated tables in the *Appendix 3B*. Second, ARMA models outperform the AR models. Third, for forecasting the best method (among the used seasonal adjustment methods) is X12 for both periodicities and for all forecast horizons. Although the forecast horizons of monthly and

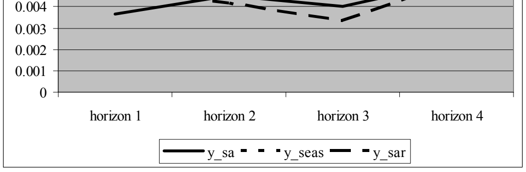
[7] I am searching only for the model with the lowest RMSFE. Such a model is definitively among the best models for forecasting, however by now I have not done any statistical test (as proposed by Diebold and Mariano (1995), Harvey et al (1997)) to check whether this model is also statistically better for forecasting than the other models with the higher RMSFE.

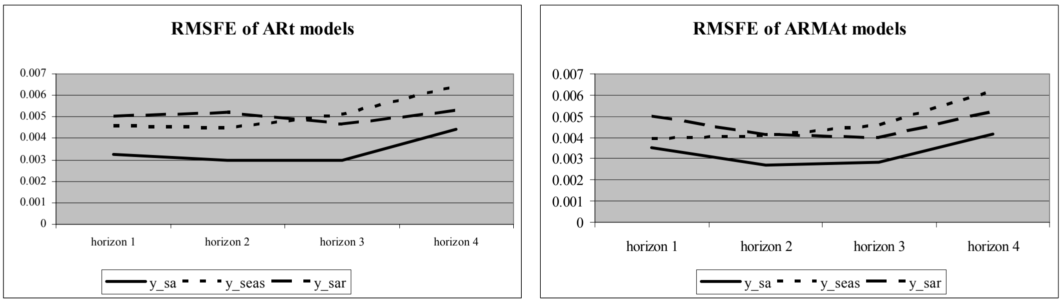
quarterly data are not directly comparable, it seems that the models with the lowest RMSFE (in the table denoted in bold) in the second and third forecast horizon of quarterly data display better forecast ability than the models with the lowest RMSFE of monthly data in the comparable forecast periods. However, the opposite is the case for the models with the lowest RMSFE for forecasting the series with the different seasonal adjustment method. As it can be figured out by comparing the tables for both periodicities in the *Appendix 3A* only for forecasting the series seasonally adjusted with X12 (with the already mentioned exception of the first and the last forecast horizon) the quarterly AR and ARMA models with trend have a lower RMSFE than the corresponding monthly AR and ARMA models. Fourth, as expected, the short-term forecasts perform on average better than medium term forecasts. In particular for monthly data, the forecast performance markedly deteriorates with the lengthening of the forecast horizon. Fifth, the models with the lowest RMSFE often switch between different forecast horizons. However, inspecting both tables in the *Appendix 3A* reveals that for AR models with and without trend this switching is less frequent and the models in many cases keep being the same for at least two forecast horizons. Sixth, within the same forecast horizon the selection of the models with the lowest RMSFE depends on the way the series is seasonally adjusted. The only exceptions is the ARMA(2,4,t) for the 4th forecast horizon of quarterly data and ARMA(2,2,t) for the 1st forecast horizon of monthly data. However, a high level of aggregation of the results in the summary tables 3i and 3ii hides the information available from the more desegregated tables in the *Appendix 3A*. They show that for AR models with and without trend the models with the lowest RMSFE, regardless of the method of seasonal adjustment, tend to be the same within the forecast horizons. Seventh, the selected models with the lowest RMSFE for certain comparable forecast horizons are only in few cases equal for both periodicities, like e.g. the ARMA(2,3,t) model for the 1st forecast horizon or ARMA(2,4,t) model for the 4th forecast horizon.

The Figures 7i – 7viii try to highlight further the importance of choosing an appropriate method of seasonal adjustment for the purposes of forecasting. For each of the 4 groups of the models the RMSFEs of the models with the lowest RMSFE are depicted for all forecast horizons and for both periodicities. Considering the monthly data for the ARMA models with and without trend as well as the AR models with trend the best method of removing a season is the X12 method for all forecast horizons. For AR models without a trend the best method of removing a season is the multiplicative method except for the 1st forecast horizon. AR and ARMA specifications without a trend display the worst forecast performance whenever the series is seasonally adjusted with seasonal dummies while for AR and ARMA models with trend the multiplicative method should definitely not be used. The similar findings hold also for the quarterly data, as only for AR and ARMA models without a trend the best method of removing a season seems to be the multiplicative method while for AR and ARMA models with trend the X12 method remains the best one. In the AR specification with 12 AR lags X12 method is superior in the models with and without trend as it can be seen from the graphs in the *Appendix 3C*.

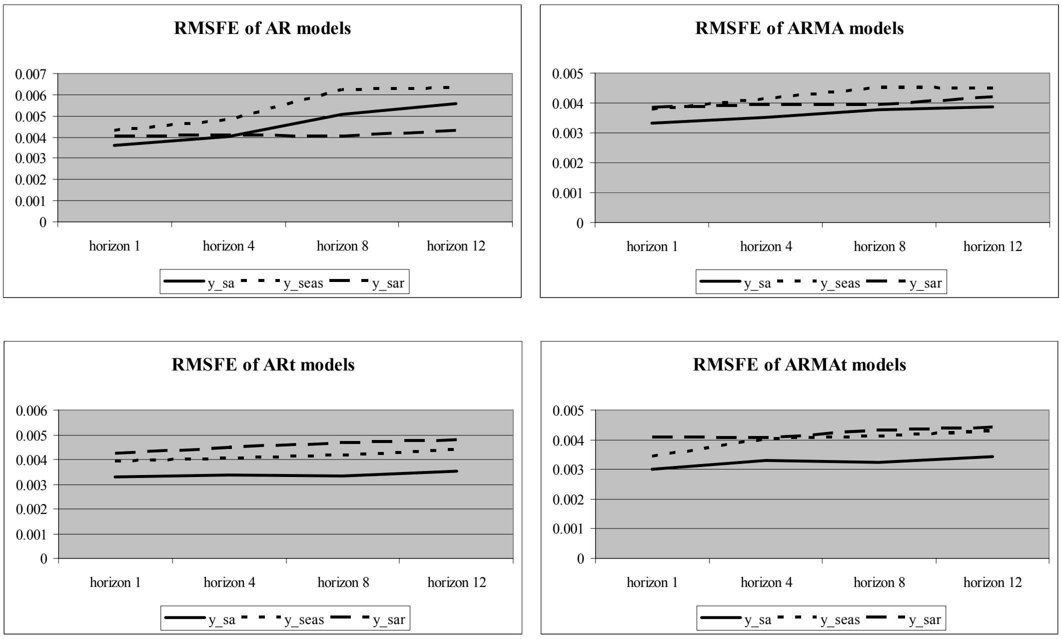
Figures 7i – 7iv: Quarterly data







Figures 7v – 7viii: Monthly data



y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression
y_sar denotes the period-by-period CPI seasonally adjusted with the multiplicative method

Do the models with the superior forecast performance deliver an optimal forecast?

In the second exercise I want to check whether the forecasts of the models with the lowest RMSFE or the superior forecast performance satisfy the list of criteria suggested by Diebold and Lopez (1996) that qualify for an optimal forecast. These criteria are the unbiasedness and the efficiency of the forecast, uncorrelated forecast residuals above their forecast horizon and the normal distribution of the residuals. The models I choose for an evaluation are the models with the lowest RMSFE (denoted in bold in the aggregation tables 3i and 3ii) for both periodicities. The results, mainly for monthly data are presented in the *Appendix 4*.

In order to check for the unbiasedness and the efficiency of the forecast, the forecast error of the certain forecast horizon is regressed on the constant and the forecast of the respective forecast horizon. An insignificant coefficient of a constant indicates very likely an unbiased forecast while an insignificant coefficient on the forecast term indicates an efficient forecast. An application of this regression on our sample of forecast errors and forecasts indicates that the forecasts for all forecast horizons for both periodicities are likely to be unbiased and efficient. The results can be obtained from the author upon request.

A further criterion for an optimal forecast relates to the forecast errors of a certain forecast horizon that should not display any autocorrelation beyond their step ahead forecast horizon reduced by one. Specifically this means that the forecast errors produced e.g. with a 4-step-ahead forecast should not display any significant autocorrelation at any lag greater than 3. Inspecting the autocorrelation functions of the forecast errors of monthly data in the *Appendix 4A* reveals the violation of this criterion only for the 1-step-ahead forecast errors while no violation of this criterion is detected for quarterly data. The results for quarterly data can be obtained from the author upon request.

The next criterion checks the normality of the forecast errors using Jarque-Bera test. Jarque-Bera test of normality is an asymptotic test and has a limited power in such a small samples as in our case of 15 observations for quarterly and 43 observations for monthly data, respectively. Due to larger number of observations for monthly data, Jarque-Bera test has been executed only for this periodicity. The results of the test together with the distributions of the forecast errors are presented in the *Appendix 4B*. The null hypothesis that the forecast errors are normally distributed is rejected only for the second forecast horizon, indicating the leptokurtic (slim or long tailed) distribution of forecast errors.

Mis-specification (in-sample) test for the models with the superior forecast performance

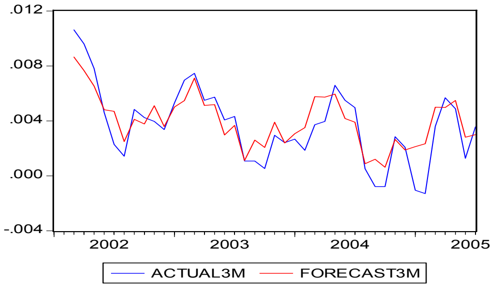
In the third exercise I want to check whether the models with the lowest RMSFE could be mis-specified in the in-sample estimation. These models are denoted in bold in the aggregation tables 3i and 3ii for both periodicities. The way I use to assess the possibility of in-sample mis-specification of the models is to check for the existence of the autocorrelated residuals of the respective models over all recursive estimation samples. In particular, I check for all recursive estimation samples the value of the Q-statistics for each lag of residuals up to 12. The main finding is that most of the models do have autocorrelated residuals in some recursive estimation samples. The only exception is the ARMA(3,1,t) model that shows an absence of the autocorrelation over all 43 recursive estimation samples. This result shows that DGP indeed differs between estimation and forecast periods. Therefore, the conclusion of our paper is that the mis-specification testing does not bring much value added for selecting the models for forecasting in a sample that is either short or subject to shifts in data generating process.

The comparison of the period-by-period CPI and forecasts over time

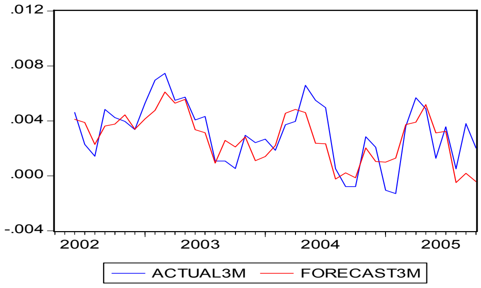
Finally, I present visually how well over time the forecasts of the models with the lowest RMSFE (denoted in bold in the aggregation tables 3ii) for 1, 4, 8 and 12 steps-ahead resemble to the actual values of m-o-m CPI. Due to the considerable short-run variation in both actual values and forecasts (as shown in the *Appendix 4C*) I consider in the Figures 8i - 8iv the 3 months moving average of both variables. Inspecting the graphs reveal that the periods of good forecasts were followed by periods of worse forecasts. In particular, forecasts tend to be relatively close to the actual value before 2004 for all forecast horizons with a marked deterioration thereafter. Similar conclusion can be derived from the graphs comparing forecasts of the models with the lowest RMSFE (denoted in bold in the aggregation tables 3i) and corresponding actual values for quarterly periodicity as displayed in the *Appendix 4D*.

Figures 8i - 8iv: Monthly data

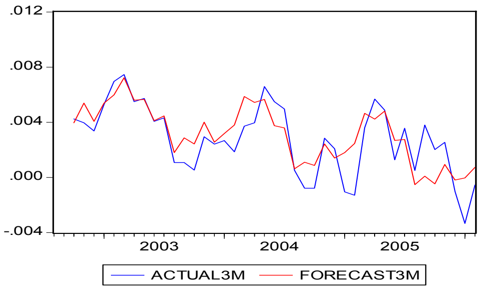
Actual values and 1-step-ahead forecast ARMA(2,2,t)



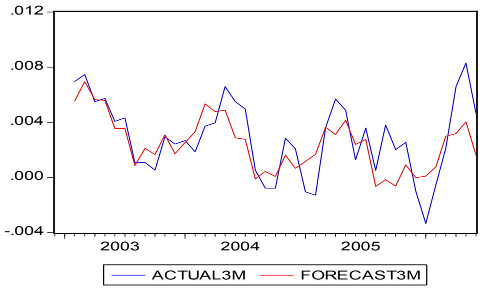
Actual values and 4-step-ahead forecast ARMA(2,3,t)



Actual values and 8-step-ahead forecast ARMA(4,4,t)



Actual values and 12-step-ahead forecast ARMA(3,1,t)



6. CONCLUSION

The forecast competition exercise has delivered a winning in-sample combination for forecasting regardless of periodicity. For forecasting period-by-period CPI the preferred method of seasonal adjustment is the X12 method. Whenever this method is used the models with trend display the best forecast performance. An additional finding is that ARMA models tend to outperform AR models. Therefore for purpose of forecasting ARMA models with trend should be used to forecast period-by-period CPI seasonally adjusted with X12 method. In this setup, a direct comparison of the models selection criteria in terms of their forecast performance provides some evidence in favor of AIC. Since ARMA models with the preferred forecast performance change rather frequently between the forecast horizons there is no receipt for a certain ARMA model specification that could be mechanically applied by the forecaster.

An important finding of this paper is that the in-sample model selection criteria on average might not be the most appropriate for a selection of the best model for forecasting in the short sample or the sample subject to structural shocks and shifts. Even for the models with trend included in a regression, there exists a relatively low correlation between the in-sample values of model selection criteria and out-of-sample performance of models. This result most likely reflects the existence of a different data generating process (DGP) in the estimation and forecast periods during the disinflation process in Slovenia. However, if any of the model selection criteria should be considered in those models than this task should be entrusted to the AIC criterion in the case of monthly data and to the R2a in the case of quarterly data.

Two main examination tests of the properties of the models with a superior forecast performance show that the models fail mis-specification check in many in-sample recursive estimation samples, yet deliver an unbiased and efficient out-of-sample forecast. The first test examines the in-sample properties of the models, in particular the existence of autocorrelation over the in-sample recursive estimation samples. The results show that autocorrelation indeed can be detected in many recursive estimation samples. Therefore this paper provides support for the finding in the literature that the mis-specification testing may not bring much of value added for the selection of the models for forecasting. The second test checks the ability of the models with the lowest RMSFE to deliver an optimal forecast. It is found that forecasts are unbiased and efficient over all forecast horizons. In addition, in most of the cases the forecast residuals are uncorrelated above their forecast horizon and normally distributed.

The examination of forecasts and period-by-period CPI over time shows that the periods of good forecasts were followed by periods of worse forecasts. In particular, forecasts tend to be relatively close to the actual value before 2004 for all forecast horizons, with a marked deterioration thereafter. The actual movement of the period-by-period CPI and of the forecasts for all forecast horizons also confirms expectations in the theory of economic forecasting that by using ARMA models short term forecasting outperforms medium term forecasting.

This work can be extended in many directions. The statistical tests as suggested by Diebold and Mariano (1995) and Harvey et al (1997) can help identifying not only the models and criteria with the lowest RMSFE but also the best criteria and the best models for forecasting. It may be the case that the other models and criteria with the higher RSMFE also qualify among the best models and criteria. Note that one could also examine confidence intervals, which however is beyond the scope of this paper. The traditional theory suggests that the calculated confidence interval around a forecast provides a good description of the likely variation in the forecast error. This theorem would probably fail in the setup where the data generating process in the estimation and forecast sample differ. An additional issue is also how to use this research that gives the results for a period of disinflation, for the forecasting of inflation in periods of a more stabilized inflation around a certain level.

REFERENCES:

Clements P. Michael and David F. Hendry (2001), *"Economic Forecasting: some lessons from recent research",* ECB Working Paper No. 82

Clements P. Michael and David F. Hendry (1999), *"On winning forecasting competitions in economics"*, Spanish Economic Review, 1, 123-160

Clements P. Michael and David F. Hendry (1998), *"Forecasting economic processes"*, International Journal of Forecasting, 14, 111-131

Diebold X. Francis and Lutz Kilian (1999), *"Unit root tests are useful for selecting forecasting models",* NBER Working Paper Series No. 6928

Diebold X. Francis and Roberto S. Mariano (1994), *"Comparing predictive accuracy"*, Technical Working Paper Series No. 169

Diebold X. Francis and Jose A. Lopez (1996), *"Forecast evaluation and combination",* Technical Working Paper Series No. 192

Enders Walter (1995), *"Applied Econometric Time Series"*, John Wiley & Sons, Inc.

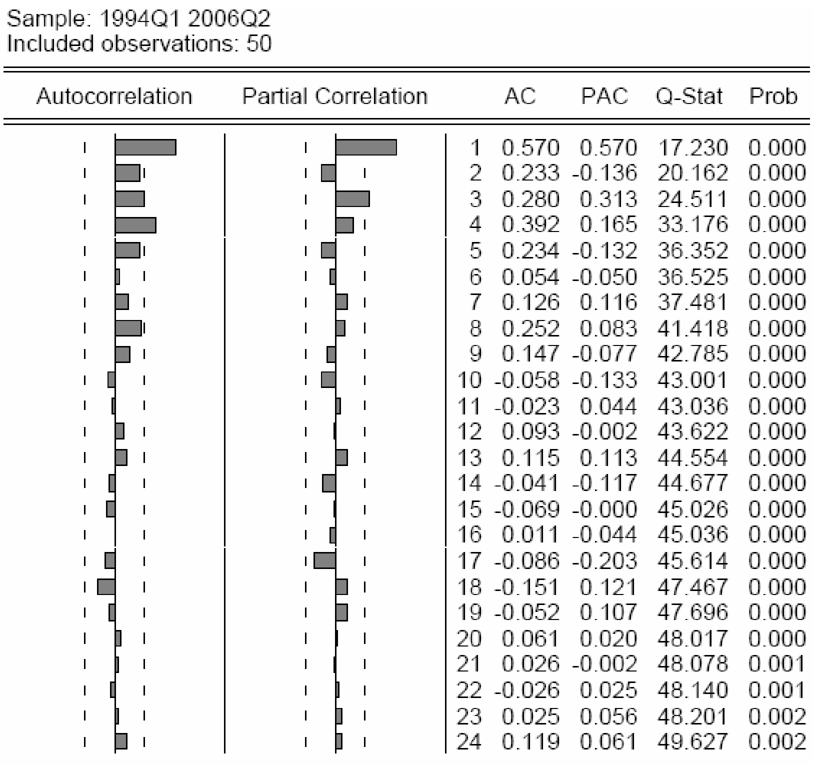
Ericsson R. Neil and David F. Hendry (2001), *"Understanding Economic Forecasts"*, The MIT Press

Hubrich Kirstin (2005), *"Forecasting euro area inflation: Does aggregating forecasts by HICP components improve forecast accuracy?"*, International Journal of Forecasting, 21, 119-136

Johnson Jack and John Dinardo (1997), *"Econometric Methods"*, The McGraw - Hill Companies, Inc.

Appendix 1:

A. Correlogram of the q-o-q CPI



Sample: 1994Q1 2006Q2
Included observations: 50

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.570	0.570	17.230	0.000
		2	0.233	-0.136	20.162	0.000
		3	0.280	0.313	24.511	0.000
		4	0.392	0.165	33.176	0.000
		5	0.234	-0.132	36.352	0.000
		6	0.054	-0.050	36.525	0.000
		7	0.126	0.116	37.481	0.000
		8	0.252	0.083	41.418	0.000
		9	0.147	-0.077	42.785	0.000
		10	-0.058	-0.133	43.001	0.000
		11	-0.023	0.044	43.036	0.000
		12	0.093	-0.002	43.622	0.000
		13	0.115	0.113	44.554	0.000
		14	-0.041	-0.117	44.677	0.000
		15	-0.069	-0.000	45.026	0.000
		16	0.011	-0.044	45.036	0.000
		17	-0.086	-0.203	45.614	0.000
		18	-0.151	0.121	47.467	0.000
		19	-0.052	0.107	47.696	0.000
		20	0.061	0.020	48.017	0.000
		21	0.026	-0.002	48.078	0.001
		22	-0.026	0.025	48.140	0.001
		23	0.025	0.056	48.201	0.002
		24	0.119	0.061	49.627	0.002

B. *Table 1:* ADF test of the q-o-q CPI and the m-o-m CPI seasonally adjusted with the X12 method:

Series	Period	None-zero mean		Non-zero mean + trend	
		ADF	lags (AIC, SBC)	ADF	lags (AIC, SBC)
q-o-q CPI	1994Q1-2006Q2	-3.6798***	3(3,0)	-4.0554***	0(0,0)
m-o-m CPI	1994M1-2006M6	-3.7994***	2(2,0)	-5.1070***	2(2,0)

*** indicate a rejection of the null hypothesis of the unit root at the 1% level.

The number of lags outside parenthesis is the actually used number of lags in the test; the numbers in parenthesis are the lags suggested by AIC and SBC, respectively.
For quarterly data the number of observations is at the limit that makes sense for carrying out ADF test.

C. Structural break (modeling a dummy variable)

A dummy variable (DUM) can enter the regression equation in the multiplicative form or the additive form. The former introduction of a dummy variable enables to differentiate between the slope coefficients of the two periods, while the latter enables to distinguish between the intercepts of the two periods. Regression results of the following regression:
$dlog(CPI) = \alpha_1 + \alpha_2 DUM_t + \beta_1 t + \beta_2 (DUM_t{*}t) + u_t$

indicate a statistically insignificant coefficient of the differential intercept α_2 as well as an insignificant differential slope coefficient β_2 for both, monthly and quarterly periodicity. Due to the possible multicolinearity between the regressors I consider two regressions separately, one with a dummy variable in the additive form and another with a dummy variable in the multiplicative form. Both regressions for monthly as well as quarterly periodicity reveal statistically significant coefficients of the dummy variable, thus indicating that the structural break in 1999 could be explained either as the upward shift in the trend before the breaking time or as the shift in the slope of the trend. Due to the introduction of the value added tax (VAT) in 1999 and the continuation of the disinflation process thereafter, I chose for modeling the dummy variable in the additive form. An approximation of the coefficient of the dummy variable in an additive form could also be obtained by subtracting the fitted value of the first observation of the trend calculated for the sole period after the occurrence of the break (1999q4-2006q2) from the fitted value of the last observation of the trend calculated for the sole period before the break (1994q1-1999q3).

Regressions for the q-o-q CPI as dependant variable

$dlog(CPI) = \alpha_1 + \alpha_2 DUM_t + \beta_1 t + \beta_2 (DUM_t{*}t) + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.044186	0.004611	9.583795	0.0000
DUM	-0.002313	0.010116	-0.228635	0.8202
@TREND	-0.001300	0.000281	-4.626092	0.0000
DUM*@TREND	0.000595	0.000358	1.663308	0.1031

$dlog(CPI) = \alpha_1 + \alpha_2 DUM_t + \beta_1 t + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.038672	0.003264	11.84861	0.0000
DUM	0.012286	0.005124	2.397916	0.0205
@TREND	-0.000933	0.000177	-5.271676	0.0000

$dlog(CPI) = \alpha_1 + \beta_1 t + \beta_2 (DUM_t{*}t) + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.043706	0.004062	10.75905	0.0000
@TREND	-0.001274	0.000253	-5.035473	0.0000
DUM*@TREND	0.000524	0.000176	2.976421	0.0046

Regressions for the m-o-m CPI as dependant variable

$dlog(CPI) = \alpha_1 + \alpha_2 DUM_t + \beta_1 t + \beta_2 (DUM_t{*}t) + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014102	0.001488	9.473785	0.0000
DUM	0.000174	0.003204	0.054343	0.9567
@TREND	-0.000131	2.97E-05	-4.418547	0.0000
DUM*@TREND	4.98E-05	3.75E-05	1.327360	0.1864

$dlog(CPI) = \alpha_1 + \alpha_2 DUM_t + \beta_1 t + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.012666	0.001026	12.35078	0.0000
DUM	0.003868	0.001592	2.430052	0.0163
@TREND	-0.000100	1.82E-05	-5.497627	0.0000

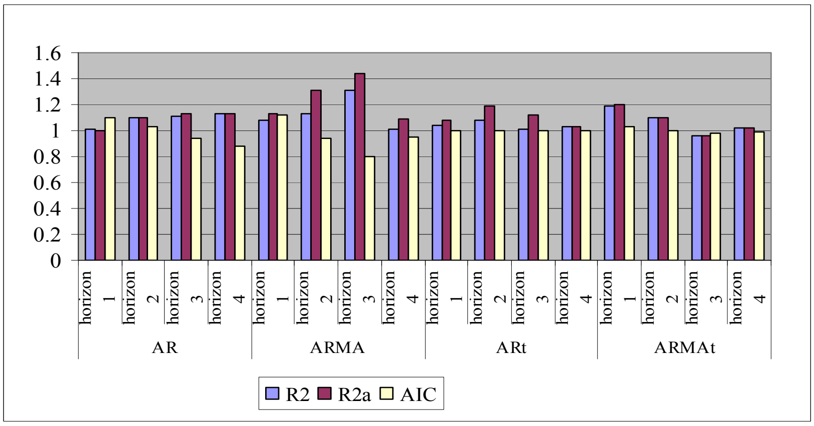
$dlog(CPI) = \alpha_1 + \beta_1 t + \beta_2 (DUM_t{*}t) + u_t$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.014139	0.001314	10.76328	0.0000
@TREND	-0.000132	2.68E-05	-4.926968	0.0000
DUM*@TREND	5.16E-05	1.85E-05	2.783285	0.0061

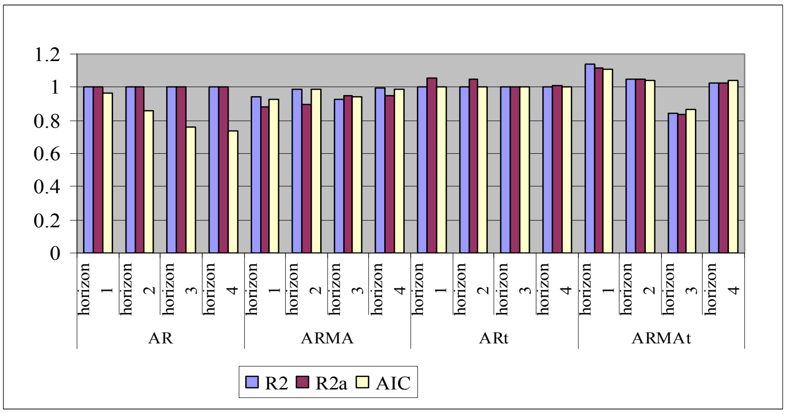
Appendix 2:

A. *Figures 2i – 2iii:* A comparison of model selection criteria in terms of their out-of-sample forecast performance, quarterly data

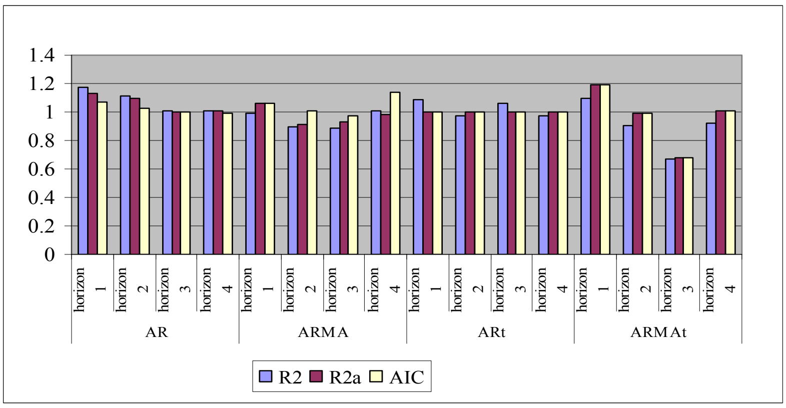
A comparison of model selection criteria for forecasting q-o-q CPI seasonally adjusted with X12



A comparison of model selection criteria for forecasting q-o-q CPI seasonally adjusted with seasonal dummies

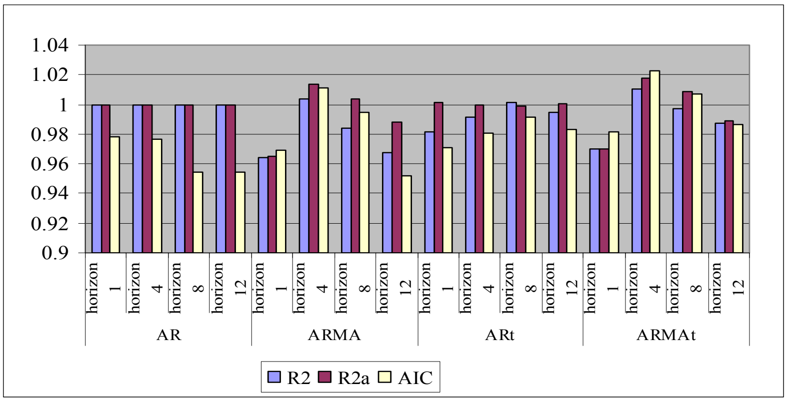


A comparison of model selection criteria for forecasting q-o-q CPI seasonally adjusted with multiplicative method

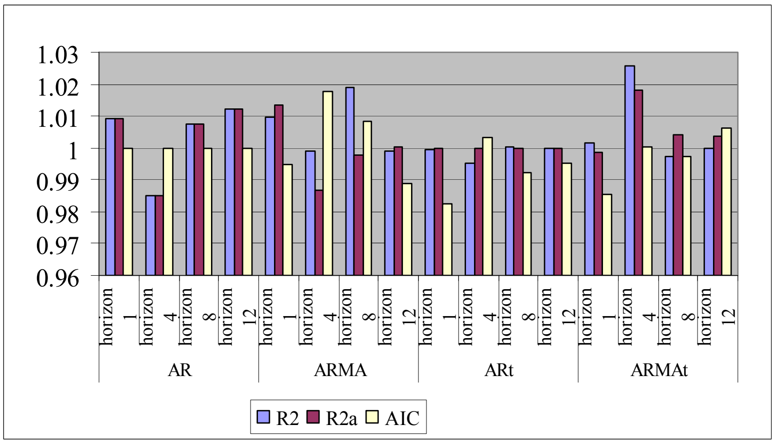


Figures 2iv – 2vi: A Comparison of model selection criteria in terms of their out-of-sample forecast performance, monthly data

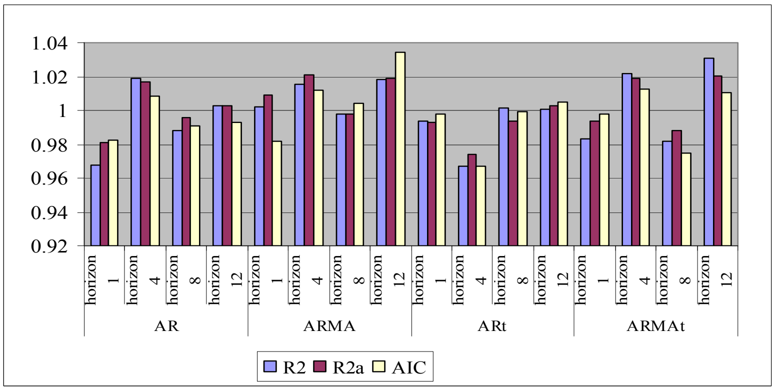
A comparison of model selection criteria for forecasting m-o-m CPI seasonally adjusted with X12



A comparison of model selection criteria for forecasting m-o-m CPI seasonally adjusted with seasonal dummies

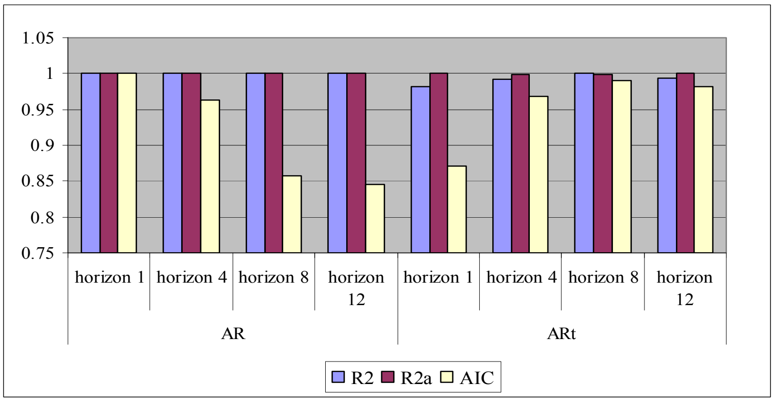


A comparison of model selection criteria for forecasting m-o-m CPI seasonally adjusted with multiplicative method

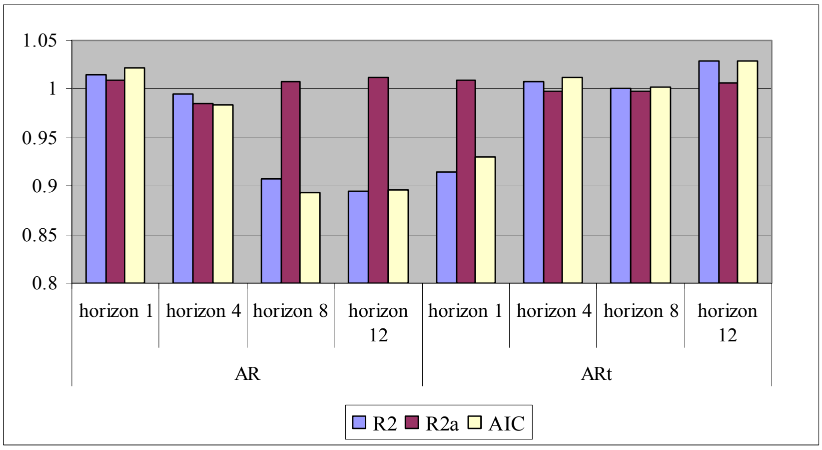


Figures 2vii – 2viii: A Comparison of model selection criteria in terms of their out-of-sample forecast performance, monthly data, models with 12 AR lags

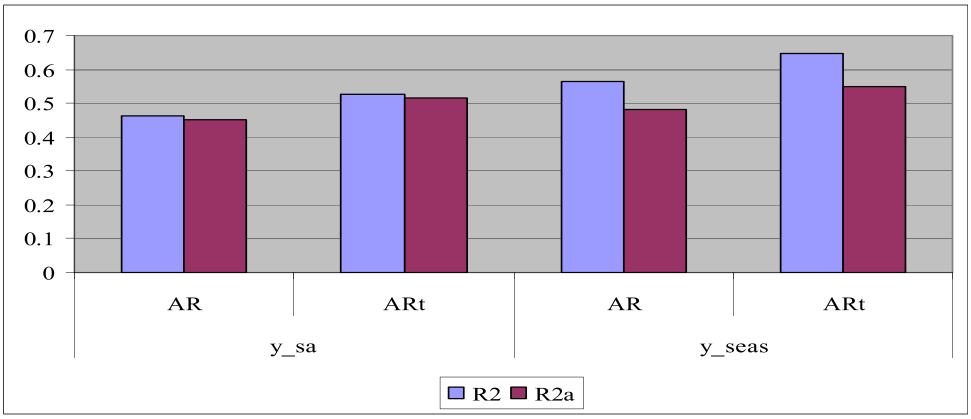
A comparison of model selection criteria for forecasting m-o-m CPI seasonally adjusted with X12

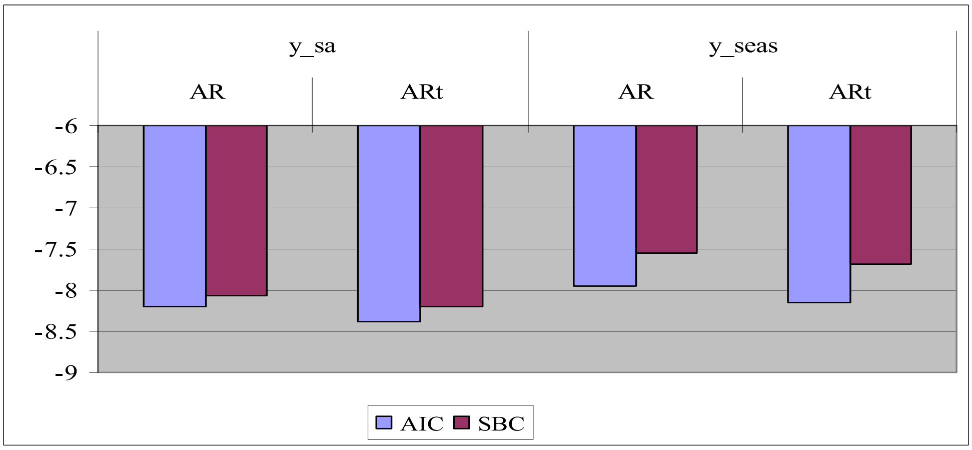


A comparison of model selection criteria for forecasting m-o-m CPI seasonally adjusted with seasonal dummies



B. *Figures 2ix – 2x:* The in-sample behavior of model selection criteria, models with 12 AR lags





Appendix 3:

A. *Table 3i*: Disaggregated table of the models with the lowest RMSFE, quarterly data

		AR		AR(t)		ARMA		ARMA(t)	
		model	RMSFE	model	RMSFE	model	RMSFE	model	RMSFE
1st horizon	y_sa	ar(2)	0.005134	ar(4,t)	**0.003235**	arma(3,1)	**0.003661**	arma(2,1,t)	**0.003536**
	y_seas	ar(4)	0.006761	ar(4,t)	0.004560	arma(4,1)	0.004924	arma(2,1,t)	0.003917
	y_sar	ar(1)	**0.004643**	ar(2,t)	0.004993	arma(2,1)	0.004736	arma(3,1,t)	0.004987
2nd horizon	y_sa	ar(2)	0.007399	ar(4,t)	**0.003001**	arma(2,3)	0.004513	arma(2,3,t)	**0.002703**
	y_seas	ar(4)	0.008665	ar(4,t)	0.004482	arma(4,3)	0.004917	arma(1,2,t)	0.004086
	y_sar	ar(2)	**0.004852**	ar(3,t)	0.005187	arma(3,3)	**0.004137**	arma(4,4,t)	0.004105
3rd horizon	y_sa	ar(4)	0.009018	ar(3,t)	**0.002973**	arma(4,4)	0.004009	arma(1,2,t)	**0.002851**
	y_seas	ar(4)	0.009679	ar(2,t)	0.005109	arma(4,3)	0.005861	arma(4,4,t)	0.004582
	y_sar	ar(2)	**0.005077**	ar(4,t)	0.004673	arma(3,3)	**0.003343**	arma(4,2,t)	0.004002
4th horizon	y_sa	ar(4)	0.009540	ar(3,t)	**0.004434**	arma(3,1)	**0.005178**	arma(2,4,t)	**0.004162**
	y_seas	ar(4)	0.009696	ar(1,t)	0.006399	arma(4,1)	0.006223	arma(2,4,t)	0.006181
	y_sar	ar(2)	**0.005560**	ar(4,t)	0.005293	arma(3,1)	0.005183	arma(4,2,t)	0.005209

Table 3ii: Disaggregated table of the models with the lowest RMSFE, monthly data

		AR		AR(t)		ARMA		ARMA(t)	
		model	RMSFE	model	RMSFE	model	RMSFE	model	RMSFE
1st horizon	y_sa	ar(4)	**0.003607**	ar(4,t)	**0.003286**	arma(2,2)	**0.003331**	arma(2,2,t)	**0.003011**
	y_seas	ar(4)	0.004303	ar(3,t)	0.003918	arma(2,2)	0.003778	arma(2,2,t)	0.003434
	y_sar	ar(4)	0.004027	ar(4,t)	0.004241	arma(3,1)	0.003852	arma(4,1,t)	0.004087
4th horizon	y_sa	ar(4)	**0.004014**	ar(4,t)	**0.003390**	arma(3,4)	**0.003511**	arma(2,3,t)	**0.003306**
	y_seas	ar(3)	0.004812	ar(3,t)	0.004023	arma(1,2)	0.004122	arma(1,3,t)	0.004014
	y_sar	ar(1)	0.004051	ar(4,t)	0.004477	arma(4,1)	0.003931	arma(3,3,t)	0.004038
8th horizon	y_sa	ar(4)	0.005096	ar(4,t)	**0.003333**	arma(3,4)	**0.003765**	arma(4,4,t)	**0.003232**
	y_seas	ar(4)	0.006239	ar(4,t)	0.004147	arma(3,3)	0.004531	arma(3,3,t)	0.004129
	y_sar	ar(3)	**0.004036**	ar(2,t)	0.004657	arma(4,1)	0.003932	arma(4,3,t)	0.004320
12th horizon	y_sa	ar(4)	0.005593	ar(4,t)	**0.003539**	arma(3,4)	**0.003879**	arma(3,1,t)	**0.003436**
	y_seas	ar(4)	0.006327	ar(4,t)	0.004383	arma(3,4)	0.004472	arma(2,4,t)	0.004270
	y_sar	ar(2)	**0.004292**	ar(1,t)	0.004767	arma(2,3)	0.004181	arma(4,1,t)	0.004414

y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression
y_sar denotes the period-by-period CPI seasonally adjusted with the multiplicative method

B. *Table 3iii:* Summary table of the models with the lowest RMSFE, models with 12 AR lags, monthly data

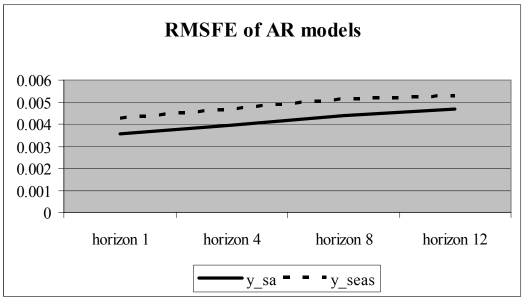
	model	RMSFE1	model	RMSFE4	model	RMSFE8	model	RMSFE12
y_sa	ar(12,t)	**0.003043**	ar(12,t)	**0.003361**	ar(11,t)	**0.003246**	ar(10,t)	**0.003488**
y_seas	ar(12,t)	0.003677	ar(3,t)	0.004023	ar(11,t)	0.003999	ar(11,t)	0.004363

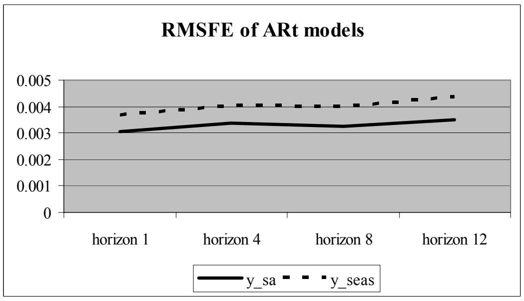
Table 3iv: Disaggregated table of the models with the lowest RMSFE, models with 12 AR lags, monthly data

		AR		AR(t)	
		model	RMSFE	model	RMSFE
1st horizon	y_sa	ar(12)	**0.003541**	ar(12,t)	**0.003043**
	y_seas	ar(11)	0.004220	ar(12,t)	0.003677
4th horizon	y_sa	ar(7)	**0.003963**	ar(12,t)	**0.003361**
	y_seas	ar(11)	0.004667	ar(3,t)	0.004023
8th horizon	y_sa	ar(11)	**0.004406**	ar(11,t)	**0.003246**
	y_seas	ar(11)	0.005145	ar(11,t)	0.003999
12th horizon	y_sa	ar(11)	**0.004704**	ar(10,t)	**0.003488**
	y_seas	ar(11)	0.005277	ar(11,t)	0.004363

y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression

C. *Figures 3i – 3ii*: A comparison of methods of seasonal adjustment, monthly data with 12 AR lags





y_sa denotes the period-by-period CPI seasonally adjusted with the X12 method
y_seas denotes the period-by-period CPI seasonally adjusted with the seasonal dummies in the regression

Appendix 4:

A. *Figures 4i – 4iv*, autocorrelation functions, monthly data

1-step ahead forecast errors
ARMA(2,2,t)

Sample: 2002M01 2005M07
Included observations: 43

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.026	-0.026	0.0315	0.859
		2	-0.442	-0.443	9.2438	0.010
		3	0.132	0.129	10.083	0.018
		4	0.137	-0.070	11.018	0.026
		5	0.041	0.197	11.104	0.049
		6	-0.026	-0.023	11.139	0.084
		7	-0.170	-0.108	12.686	0.080
		8	-0.015	-0.066	12.698	0.123
		9	0.021	-0.153	12.723	0.176
		10	-0.060	-0.059	12.935	0.227
		11	0.138	0.160	14.082	0.228
		12	0.033	0.044	14.149	0.291
		13	-0.061	0.144	14.387	0.347
		14	0.053	0.013	14.576	0.408
		15	0.105	0.114	15.331	0.428
		16	0.065	0.049	15.637	0.479
		17	0.019	0.081	15.662	0.548
		18	-0.090	-0.036	16.283	0.573
		19	-0.082	-0.086	16.824	0.602
		20	0.025	-0.072	16.876	0.661

4-step ahead forecast errors
ARMA(2,3,t)

Sample: 2002M04 2005M10
Included observations: 43

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.045	-0.045	0.0933	0.760
		2	-0.437	-0.440	9.1107	0.011
		3	-0.017	-0.081	9.1249	0.028
		4	0.106	-0.117	9.6865	0.046
		5	0.053	0.013	9.8295	0.080
		6	0.087	0.119	10.223	0.116
		7	-0.093	-0.044	10.689	0.153
		8	-0.306	-0.277	15.858	0.044
		9	-0.015	-0.174	15.871	0.070
		10	0.193	-0.125	18.062	0.054
		11	0.051	-0.048	18.221	0.077
		12	-0.141	-0.116	19.454	0.078
		13	-0.058	-0.052	19.667	0.104
		14	-0.001	-0.117	19.667	0.141
		15	0.112	-0.010	20.529	0.153
		16	0.196	0.090	23.283	0.106
		17	-0.000	0.082	23.283	0.140
		18	-0.140	0.063	24.799	0.131
		19	-0.125	-0.123	26.062	0.128
		20	0.137	0.004	27.646	0.118

8-step ahead forecast errors
ARMA(4,4,t)

Sample: 2002M08 2006M02
Included observations: 43

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.082	-0.082	0.3100	0.578
		2	-0.332	-0.341	5.5061	0.064
		3	0.029	-0.041	5.5456	0.136
		4	0.010	-0.119	5.5508	0.235
		5	0.034	0.021	5.6084	0.346
		6	-0.042	-0.081	5.7023	0.457
		7	0.061	0.082	5.9014	0.551
		8	-0.282	-0.360	10.293	0.245
		9	-0.093	-0.132	10.789	0.290
		10	0.135	-0.179	11.862	0.294
		11	0.069	-0.001	12.153	0.352
		12	0.049	-0.007	12.301	0.422
		13	-0.063	-0.000	12.559	0.482
		14	-0.084	-0.168	13.031	0.524
		15	0.099	0.073	13.701	0.548
		16	0.212	0.081	16.923	0.391
		17	0.010	0.089	16.930	0.459
		18	-0.128	-0.009	18.202	0.442
		19	-0.130	-0.096	19.562	0.421
		20	-0.020	-0.087	19.595	0.483

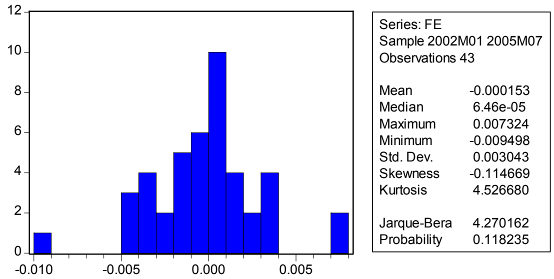
12-step ahead forecast errors
ARMA(3,1,t)

Sample: 2002M12 2006M06
Included observations: 43

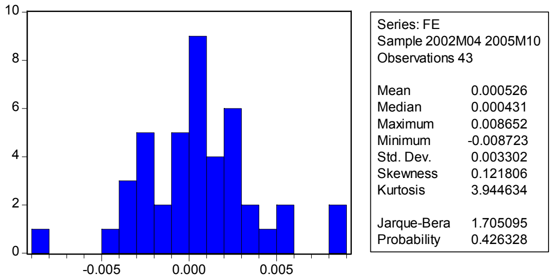
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.020	0.020	0.0183	0.892
		2	-0.338	-0.339	5.4092	0.067
		3	-0.097	-0.091	5.8605	0.119
		4	-0.094	-0.235	6.3029	0.178
		5	0.053	-0.020	6.4444	0.265
		6	-0.011	-0.167	6.4510	0.375
		7	0.145	0.153	7.5809	0.371
		8	-0.175	-0.324	9.2681	0.320
		9	-0.091	0.054	9.7434	0.372
		10	0.149	-0.084	11.046	0.354
		11	0.045	0.088	11.168	0.429
		12	0.052	-0.037	11.338	0.500
		13	0.003	0.147	11.338	0.583
		14	-0.078	-0.152	11.749	0.626
		15	-0.061	0.140	12.007	0.678
		16	0.110	0.007	12.877	0.682
		17	0.100	0.164	13.628	0.693
		18	-0.045	-0.030	13.786	0.743
		19	-0.186	-0.068	16.576	0.619
		20	-0.075	-0.095	17.043	0.650

B. *Figures 4v – 4viii*, Jarque-Bera distributions, monthly data

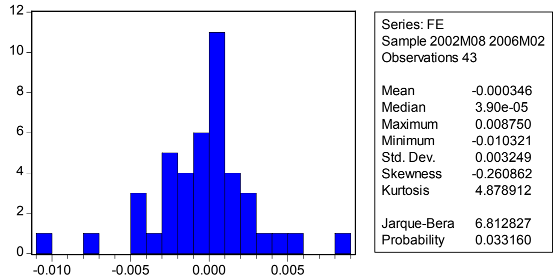
1-step ahead forecast errors
ARMA(2,2,t)



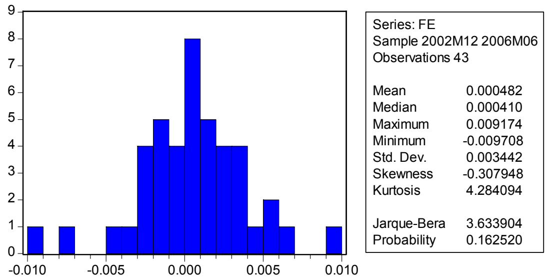
4-step ahead forecast errors
ARMA(2,3,t)



8-step ahead forecast errors
ARMA(4,4,t)

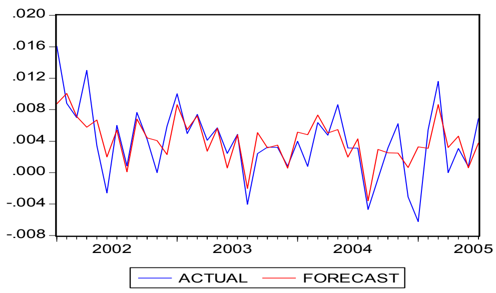


12-step ahead forecast errors
ARMA(3,1,t)

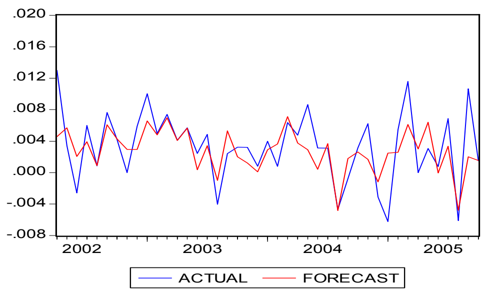


C. *Figures 4ix – 4xii*, monthly data, seasonal adjustment method is X12

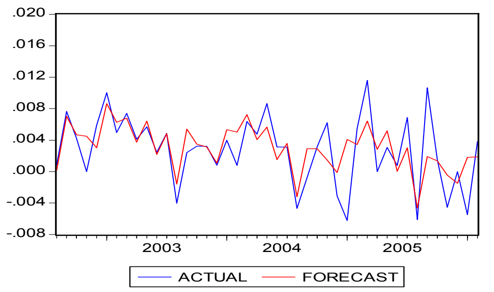
Actual values and 1-step-ahead forecast
ARMA(2,2,t)



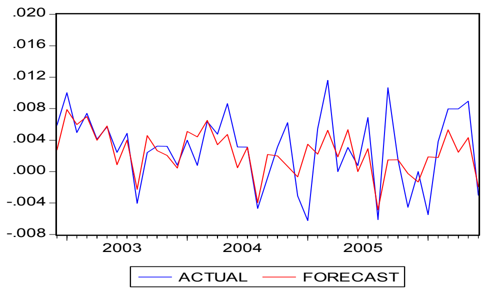
Actual values and 4-step-ahead forecast
ARMA(2,3,t)



Actual values and 8-step-ahead forecast
ARMA(4,4,t)

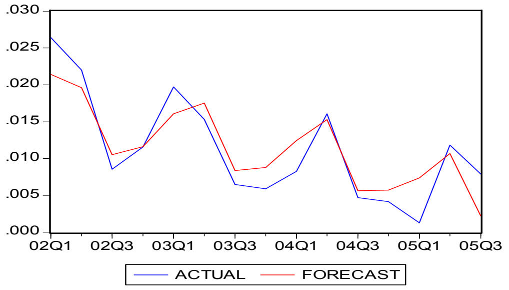


Actual values and 12-step-ahead forecast
ARMA(3,1,t)

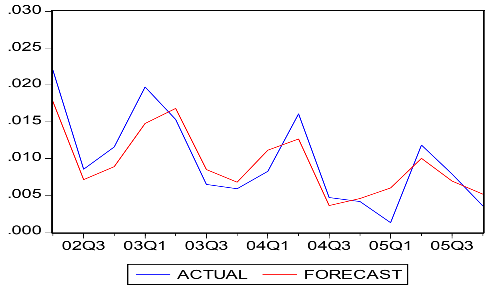


D. *Figures 4xiii – 4xvi*, quarterly data, seasonal adjustment method is X12

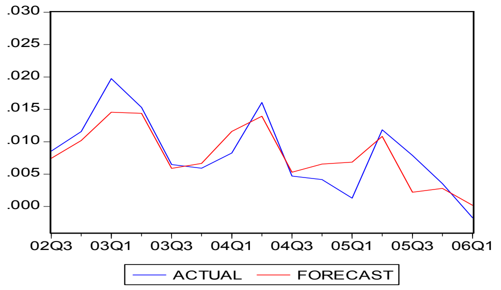
Actual values and 1-step-ahead forecast ARMA(4,0,t)



Actual values and 2-step-ahead forecast ARMA(2,3,t)



Actual values and 3-step-ahead forecast ARMA(1,2,t)



Actual values and 4-step-ahead forecast ARMA(2,4,t)