Forecasting Brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon

Thiago C. Carlos
Emerson Fernandes Marçal
Forecasting Brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon

Thiago C. Carlos
UBS Brazil Broker Dealer

Emerson Fernandes Marçal
CEMAP-EESP-FGV and CCSA-Mackenzie
Rua Itapeva 474, sala 1012
ZIP code: 01332-000
São Paulo-SP, Brazil
Email: efmarcal@gmail.com.

November, 2013

ABSTRACT

This work aims to compare the forecast efficiency of different types of methodologies applied to Brazilian Consumer inflation (IPCA). We will compare forecasting models using disaggregated and aggregated data over twelve months ahead. The disaggregated models were estimated by SARIMA and will have different levels of disaggregation. Aggregated models will be estimated by time series techniques such as SARIMA, state-space structural models and Markov-switching. The forecasting accuracy comparison will be made by the selection model procedure known as Model Confidence Set and by Diebold-Mariano procedure. We were able to find evidence of forecast accuracy gains in models using more disaggregated data.

Keywords: Inflation, forecasting, ARIMA, space-state model, Markov-switching, Model Confidence Set.

JEL Codes: C53; E31; C52

1 Corresponding author: efmarcal@gmail.com.
Forecasting Brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon

ABSTRACT

This work aims to compare the forecast efficiency of different types of methodologies applied to Brazilian Consumer inflation (IPCA). We will compare forecasting models using disaggregated and aggregated data over twelve months ahead. The disaggregated models were estimated by SARIMA and will have different levels of disaggregation. Aggregated models will be estimated by time series techniques such as SARIMA, state-space structural models and Markov-switching. The forecasting accuracy comparison will be made by the selection model procedure known as Model Confidence Set and by Diebold-Mariano procedure. We were able to find evidence of forecast accuracy gains in models using more disaggregated data.

Keywords: Inflation, forecasting, ARIMA, space-state model, Markov-switching, Model Confidence Set.

JEL Codes: C53; E31; C52

1 INTRODUCTION

Inflation forecasting exercises play an important role within the empirical work of macroeconometrics. In the Brazilian case, the period of sharply rising inflation in the 1980s and first half of the 1990s led to the adoption of several stabilization plans, creating awareness of the benefits of price stability, which finally occurred with the Real Plan in 1994. Economic stabilization in recent historical terms also creates problems for the analysis of inflation dynamics in Brazil due to a relatively short time series, a common problem in developing countries. Both the problems of data availability and of abundance of institutional changes make it difficult to project the future through the past information, as Schwartzman (2006) observes. A second important step in economic stabilization came from the adoption of an inflation-
targeting regime in June 1999. The adoption of inflation targeting became a fundamental difference in the development of econometric inflation models in Brazil. According to Chauvet (2000), to make the inflation-targeting regime operational, the comparison of inflation forecasts with the stated goal becomes an important tool in monetary policy decision making, given the lag of transmission mechanisms.

Inflation forecasting is not only important for the monetary authority; it is also of great importance to private agents, who try to understand and react to central bank behavior. A more accurate inflation forecast is also very useful in fiscal policy, wage negotiations, and also in the financial markets.

In this context, it is crucial to obtain inflation forecasts even more precise for several time horizons. Seeking the best forecast techniques, several approaches have been studied over time, including aggregated and disaggregated models, with linearity and nonlinearity; structural models for time series; and the Phillips curve based models, are among some of the techniques studied.

In this work we seek to identify the best estimation technique of Brazilian inflation among the univariate time-series models. We use linear and non-linear models and aggregated and disaggregated data, evaluating out-of-sample forecast performance for periods of 1-12 months ahead using techniques such as Diebold and Mariano (1995) and the Model Confidence Set by Hansen et al. (2010).

The work is divided into five parts, besides this introduction. The following section is a literature review of inflation forecasting methodologies and model comparison methods. The third section will provide a brief description of the methodology and models that will be estimated, as well as a breakdown of the Model Confidence Set. It will also describe how the data were treated, their peculiarities, and
the breakdowns used. The fourth section will present the results of the estimated models and comparisons of models. In the fifth section we discuss the limitations and possible extensions of the work. Finally, the conclusions will be presented.

2 LITERATURE REVIEW

2.1 FORECASTING METHODS

One of the possibilities that could result in more accurate forecasts involves breakdown of the data. This alternative focuses on the decomposition of the index into several subcomponents, each with its associated weight; instead of forecasting the main variable, each of the subcomponents is projected individually, to be reaggregated later. In this approach, the breakdown of the data can increase the accuracy of the forecast, insofar as each of its subcomponents may be modeled by their individual characteristics. Lütkepohl (1984) argued that if the data generating process is known, it is preferable to design multiple disaggregated time series and then aggregate them, rather than forecasting the aggregated series directly. However, in practice, the data generating process is unknown, and Lütkepohl presents evidence that given the variability of the model specification and estimation, it may be preferable to design the aggregate variable directly. The loss of information due to aggregate series is also addressed by Khon (1982), Palm and Nijman (1984), Rose (1977), and Tiao and Guttman (1980). It is not clear whether the aggregation of forecasts actually improves their accuracy, as the result also depends on the level of disaggregation of the series. According to Hubrich and Hendry (2006), the potential specification error in the forecast model with disaggregated variables is due to the selection of models, uncertainty in the estimation,
as well as measurement errors and structural breaks affects the forecast result. This explains why the theoretical results of the expected forecast are not confirmed in empirical studies. For example, Hubrich (2005) indicated that the aggregation of the forecasts of the components of inflation does not necessarily help with 12-month forecasts for the Eurozone. Duarte and Rua (2007) found an indirect relationship between forecast horizon and amount of information, i.e. more disaggregated models are more efficient in short-term forecasts, while aggregate models are more efficient for longer horizons.

Another way to incorporate richer structures for modeling economic series is the use of forecast models of aggregate series with nonlinear structure. Economic series, in general, are subject to changes in economic policies, a recurring problem in emerging countries, which may generate changes of regime. Possible structural changes in economic series have motivated the development of models that incorporate this particular feature. Many models of non-linear time series have been proposed in the literature, such as the bilinear model (Granger and Anderson, 1978); threshold autoregressive (TAR) model (Tong, 1978); the state-space model used in engineering to represent a variety of physical processes, and its introduction and comparison of applications in engineering and econometrics performed by Mehra (1974); and the Markov-switching model proposed by Hamilton (1989).

A large number of studies on nonlinearity in inflation investigate Friedman's hypothesis that high inflation rates lead to a higher variance of future inflation. Friedman suggests that uncertainty about the inflation regime may be the fundamental source of the positive relationship between inflation and volatility. The Markov-switching models have been especially designed to capture this kind of phenomenon in
the series. Studies in this regard are found in Evans and Wachtel (1993), Kim (1993), Bidarkota and McCulloch (1998), and Bidarkota (2001).

Another alternative studied involves a combination of the models forecasts, so that a forecast derived from the combination of several other forecasts may result in greater accuracy in the estimation. Various combinations are possible, but in many cases, a simple average of the forecasts can significantly improve the results, as noted by Clemen (1989). The first to study the combination of forecasts on economic situations were Reid (1968, 1969) and Bates and Granger (1969). Earlier studies in forecast combination were founded in the fields of statistics and psychology. To go deeper on the subject, Clemen (1989) made an extensive literature review of forecast combination in his work.

In the Brazilian literature, most alternative inflation forecasting models use the Phillips curve. Bogdanski et al. (2000) presented a small size structural model to represent the monetary policy transmission channels, estimating a Phillips curve as part of the model. In an attempt to build a leading indicator to anticipate Brazilian inflation, Chauvet (2000) created a dynamic factor model using the Kalman Filter. Among the attempts to estimate the Phillips curve for Brazil, Schuwartzman (2006) estimated a disaggregated Phillips curve, using the three-stages least square method, using quarterly data for different samples. Areosa and Medeiros (2007) derived a structural model for inflation in an open economy, using a representation of the neo-Keynesian Phillips curve and a hybrid curve. Sachsida et al. (2009) estimated a Phillips curve with quarterly data, in which they adopted a Markov-switching model, noting that the estimated coefficients are highly sensitive, and suggesting the inadequacy of the Phillips curve for explaining inflation dynamics in the Brazilian economy. Arruda, Ferreira and
Castelar (2011) compared inflation forecasts in Brazil with monthly data using linear and nonlinear time series and the Phillips curve.

2.2 COMPARISON OF PREDICTIVE POWER OF MODELS

Not only are estimation techniques important, but so are comparisons of results from different forecast methods. Several studies in the literature have addressed the problem of selecting the best model from a set of forecasting models. It is well known that increasing the number of variables in a regression improves the model adhesion within the sample—reducing, for example, the mean squared error—but the worst forecasts end up being from those models with the greatest number of variables. Engle and Brown (1985) compared the model selection procedures based on six information criteria with two testing procedures, showing that the criteria that more heavily penalizes the overparametrization of the models showed the best performance. Inoue and Kilian (2006) compared the selection procedures based on out-of-sample information criteria and evaluation by the mean squared forecast error. A similar problem, known as multi comparison with control, occurs when all objects are compared to a benchmark model, i.e. when a base model is chosen whose performance is compared with all other models. Along these lines, a test for equal predictive ability was proposed by Diebold and Marian (1995), consisting of a formal test between two competing forecast models under the null hypothesis of equivalence of accuracy. Harvey, Leybourne, and Newbold (1997) suggested a modification to Diebold and Mariano’s test to best treat cases of small samples. Instead of testing equivalent forecast accuracy, White (2000) proposed a procedure to test the superior predictive ability (SPA) of the models. While the equivalence test has a single null hypothesis to be
tested, the SPA test has multiple hypotheses. Hansen (2005) presented a new SPA test, following White’s construction, but with a different statistic test, with a distribution dependent on the sample for the test of the null hypothesis. Finally, Hansen, Lunde and Nason (2010) introduced the Model Confidence Set (MCS), which is a multi-model comparison procedure without the need to set a benchmark model. The MCS is a methodology that determines the "best" model or set of "best" models within a confidence interval and has several advantages over previously proposed methods.

3 MODELLING AND FORECASTS COMPARISON TECHNIQUES

3.1 MODEL CONFIDENCE SET

The Model Confidence Set (MCS) is a methodology for comparing forecasts from a set of models, determining the "best" model or set of "best" models within a confidence interval. The methodology that we follow was introduced by Hansen et al. (2010).

According to Hansen, an attractive feature of the MCS is that it recognizes the limitations of the data, evaluating the sample information about the relative performance of the collection of models under consideration. If the sample is informative, the MCS will result in a single "best" model. For a sample offering less informative data, it is difficult to distinguish the models, which may result in an MCS containing several or even all models. This feature differs from the existing selection methods that choose a single model without considering the existing information in the data.
Another advantage of MCS is that it is possible to comment on the significance that is valid in the traditional manner, a property that is not satisfied in the approach typically used to report the \( p\)-value of a multiple comparison of pairs of models. Thus, the advantage of MCS is that the procedure allows more than one model to be considered the "best", resulting in a set of "best" models.

Let us suppose that for a collection of models that compete with each other, \( \mathcal{M}_0 \), the procedure determines a set, \( \mathcal{M}^* \), which consists of the best set of forecast models. The MCS procedure is based on an equivalence test, \( \delta_{\mathcal{M}} \), and a rule of elimination, \( e_{\mathcal{M}} \). The equivalence test, \( \delta_{\mathcal{M}} \), is applied to the set of objects \( \mathcal{M} = \mathcal{M}_0 \). If \( \delta_{\mathcal{M}} \) is rejected, there is evidence that the models in \( \mathcal{M} \) are not equally good, and thus the rule of elimination, \( e_{\mathcal{M}} \), is used to remove the object with the worst performance in \( \mathcal{M} \). This procedure is repeated sequentially until \( \delta_{\mathcal{M}} \) is accepted, and the MCS will be defined as the set of “survival” models. The same level of significance is applied in all tests, which asymptotically ensures a significance level of \( \alpha \), \( P(\mathcal{M}^* \subset \widehat{\mathcal{M}}^{*}_{1-\alpha}) \geq 1 - \alpha \), and in the case where \( \mathcal{M}^* \) consists of only one object we have the stronger results that \( \lim_{n \to \infty} P(\mathcal{M}^* \subset \widehat{\mathcal{M}}^{*}_{1-\alpha}) = 1 \), i.e., a single object "survives" on tests when \( n \) tends to infinity and is never removed, so its probability is equal to 1. The MCS also determines a \( p\)-value for each model (MCS \( p\)-value), which follows an intuitive form related to the \( p\)-value of the individual tests of equivalence between models. Briefly, for a given model \( i \in \mathcal{M}_0 \), the \( p\)-value of MCS, \( \widehat{p}_i \), is on the border where \( i \in \widehat{\mathcal{M}}^{*}_{1-\alpha} \), if and only if \( p_i \geq \alpha \). Thus, a model with low \( p\)-value is
unlikely to be one of the models selected for the set of best models $\mathcal{M}^*$, while a model with a high $p$-value probably will be part of $\mathcal{M}^*$.

In this sense, there is no need to assume a particular model as being the "best", nor is the null hypothesis test defined by a single model. Each model is treated equally in comparing and assessing only its out-of-sample predictive power. This is one of the most interesting features of the MCS that will be used in this work. The MCS was implemented using the OxMetrics MULCOM package of Hansen and Lunde (2010).

3.2 SARIMA MODELS

ARIMA models are a method widely discussed in the literature, initially introduced by Box and Jenkins (1976). These models are also often used for seasonal series, SARIMA, capturing the seasonal and non-seasonal dynamics of the series. The estimated models in this work, besides ARIMA seasonal components, we also allowed the existence of deterministic seasonality, by adding seasonal dummies. Thus, the model identification process would incorporate the type of seasonality that best fits the series, or even the two types. The estimated model is as follows:

$$\phi(L)\varphi(L^s)(\Delta y_t - \beta_1 D_1 - \cdots - \beta_{12} D_{12}) = \theta(L)\delta(L^s)e_t$$

(1)

These models were estimated by the X-12 ARIMA software provided by the United States Census Bureau, using the procedure of automatic model selection (autmodl) derived from that used by TRAMO (Gómez & Maravall (2001). The model allowed the automatic identification of outliers\(^3\), the transformation of the series (in logarithm or not) was also automatically defined, and the number of autoregressive or

\(^2\) Seasonal Autoregressive Integrated Moving Average Model.

\(^3\) Some models employ more rigorous criteria for identifying outliers in order to avoid excessive identification of outliers.
moving average components was limited to three, in order to avoid an excessive number of parameters to be estimated. The models were selected by the lowest value on Akaike information criteria (Akaike (1973)).

3.3 STRUCTURAL TIMES SERIES MODELS

According to Watson and Engle (1983), the use of unobservable variables in economics is widely accepted as a useful approach to describe economic phenomena. The general idea of the structural model for time series is that the series are a sum of components, not necessarily observed, such as tendency, seasonality, and cycles, wherein each component evolves according to a particular dynamic. Structural models in a state-space representations offer a very interesting approach to forecasting, by allowing parameters to be stochastic. One can define a state-space model as follows:

\[ y_t = Z_t \alpha_t + \varepsilon_t \quad \varepsilon_t \sim N(0, H_t) \]  
\[ \alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim N(0, Q_t) \]

where \( y_t \) is a vector of observations \( p \times 1 \); \( \alpha_t \) is called the state vector, is unobserved, and has dimension \( m \times 1 \); \( \varepsilon_t \) and \( \eta_t \) are independent error terms. The model estimation is done using the Kalman filter combined with maximum likelihood, in which the forecast error is minimized. The Kalman filter is composed by a set of equations to estimate recursively in time the mean and conditional variance of the state vector.

The estimated state-space structural model was decomposed into trend \( (\mu_t) \), seasonality \( (\gamma_t) \), and stochastic cycle \( (\psi_t) \). It was also included the identification of

---

4 The model specification programmed for the X-12 can be obtained upon request to thiago.carlos@gvmail.br. More details about the automatic selection models procedure in X-12 ARIMA can be seen in the reference manual available in http://www.census.gov/srd/www/x12a/.
outliers in order to minimize normality problems in the series. All models were estimated using STAMP from Oxmetrics. Thus, the estimated model follows the form of the basic structural model with cycle, which can be written as follows:

\[ y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2) \]  
\[ \mu_t = \mu_{t-1} + \beta_{t-1} + \nu_t \quad \nu_t \sim N(0, \sigma^2) \]  
\[ \beta_t = \beta_{t-1} + \eta_t \quad \eta_t \sim N(0, \sigma^2) \]  
\[ \gamma_t = -\sum_{j=1}^{z} y_{t-j} + \xi_t \quad \xi_t \sim N(0, \sigma^2) \]

where \( \psi_t \) is the stochastic cycle, which can be defined by

\[ \begin{pmatrix} \psi_t \\ \psi_i \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \text{sen} \lambda_c \\ \text{sen} \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{i-1} \end{pmatrix} + \begin{pmatrix} k_t \\ k_i \end{pmatrix} \]  

where \( k_t \) and \( k_i \) are mutually uncorrelated disturbances with common variance \( \sigma^2 \), and \( 0 \leq \rho \leq 1 \).

3.4 MARKOV SWITCHING MODEL

Markovian regime switching models are non-linear models of time series analysis in which there is a probability of transition between regimes connected to the immediately preceding period. These models are very useful in modeling series that present periods with distinct behaviors. For example, one should not expect a recessionary economy to behave in the same way as an economy in a period of expansion. In the case of inflation, Friedman's hypothesis that higher inflation rates lead to increased volatility is the best suited for the possibility of alternating regimes of inflation. This nonlinearity in economic series can be modeled from regime-switching models. We will use a Markov-switching autoregressive model, allowing two regimes, one for periods of high inflation and the other for periods of low inflation. In addition,
our model will include dummies for deterministic seasonality. It is worth noting that we allow for the existence of two regimes in the variance, according to Friedman’s hypothesis. The estimated model will have the following form:

\[
y_t = \begin{cases} 
  c_1 + \phi_1 y_{t-1} + \sum_{i=1}^{11} D_i + \varepsilon_{1t} & \text{if } s_t = 1 \\
  c_2 + \phi_2 y_{t-1} + \sum_{i=1}^{11} D_i + \varepsilon_{2t} & \text{if } s_t = 2 
\end{cases}
\]  

where \( s_t \) takes values \( \{ 1, 2 \} \), being the following transition probabilities between regimes defined:

\[
P = \begin{bmatrix} p_{11} & p_{12} \\
p_{21} & p_{22} \end{bmatrix}
\]

In which \( p_{ij} \) can be defined as the probability of the regime \( i \) being followed by the regime \( j \). The phenomenon is governed by a non-observed process in which the probabilities model the transition from one conditional function to another. The model estimation process depends on the construction of a likelihood function, and its optimization using an algorithm similar to that suggested by Hamilton (1989). First, the likelihood function is optimized, next the filtered and smoothed probabilities are calculated, and finally standard deviations and statistics are calculated for inference.

4 DATA

The series used for the models are from IPCA - the official consumer price inflation measure in Brazil - , released by the Brazilian Institute of Geography and Statistics (IBGE)—, starting in January 1996 and ending in March 2012. Currently the IPCA has the following subdivisions, in hierarchical order: Groups (9); Subgroups (19) Items (52), and Sub-items (365). However, the subdivisions of IPCA change over time,
following the disclosure of the Consumer Expenditure Survey (POF) by IBGE, to more adequately reflect the changes in consumption patterns of the target population over the years. Altogether, throughout the IPCA reporting period there were five changes in the basket of goods used, and within the review period of the study there were three changes: in August 1999, July 2006, and January 2012.  

The most recent change to the weighting structure took place in January 2012, resulting in a reduction in the number of sub-items, from 384 to the present 365, due to the inclusion of 31 new sub-items and exclusion of 50, without changes in the other subdivisions: the 9 groups and 52 items remained unchanged. However, previous alterations showed more significant changes. It was decided to make the sets of groups and items of the IPCA compatible with the more current structure, using the IBGE’s structure converter. Thus, we worked with 9 groups and 52 items in the forecast in more disaggregated levels, insofar as there were fewer changes over the period under review, making its compatibilization easier. Only two of the current items lack a structure corresponding to the period between January 1991 and July 1999, and therefore begin in August 1999. Another six items have changed more significantly, but could be reconciled to complete the series.  

Another way to analyze the IPCA in a disaggregated form is by using the item classification system of the Central Bank of Brazil (BCB), which follows international standards recommended by the United Nations (UN). To this end, the BCB announces the items that comprise the following classifications: services, durable goods, nondurable goods, semi-durable goods, tradable goods, non-tradable goods,

---


6 More details about the compatibility of the data and changes in the IPCA structures over time can be seen in the APPENDIX.
monitored prices. With the change in the POF in January 2012, the BCB also made some changes to the previous classifications, and because the objective of this work is to test the best breakdowns for projected inflation, we decided to make the forecasts in both the old and the new classifications.

The main differences between the BCB’s current and previous classification are in the groups of services and monitored prices. In the new classification, services include the subgroup "Food away from home" (weight of 7.97% of the IPCA in Jan-12) and the sub-item "Airline tickets" (weight of 0.57% of the IPCA in Jan-12). The group of monitored prices excluded the item "Airline tickets" and also the sub-item "Cell Phone" (weight of 1.52% of the IPCA in Jan-12). Thus, the new classification gives the Services group a weight of 33.72% for January 2012, whereas the old classification would have given 23.66% for the same period. Table 1 presents the changes of weighting for each instance of change in the structures of the IPCA because of POF changes. One of the classifications that draws the most attention is that of monitored prices, given the great weighting change over the period. At the beginning of the review period, in January 1996, the share of monitored prices was only 13.25%, reaching 31.29% in July 2006, but being greatly reduced in the last weighting change in January 2012 to 26.70%.

---

7 These changes are significant because the subgroup "Food away from home", despite being a service, may suffer a major influence of the variation in food prices, which are more volatile, changing the dynamics of the series. Moreover, the subsection "Air tickets," has undergone a relatively recent change in methodology for calculating (February 2010), which led to increased volatility of the sub-item, which despite the small weight because of its high variation rates month to month, may have a significant influence on the final outcome of the group.

8 The details of the BCB’s classifications and the differences with the new classification can be obtained from the Box "Updates of the Weighting Structures IPCA and INPC and of IPCA Classifications of the Quarterly Inflation Report of December 2011. Available at http://www.bcb.gov.br/htms/relinf/port/2011/12/ri201112b3p.pdf
We will thus employ the IPCA forecasts with the following breakdowns, in which we will use the names defined in Table 2. With models at the aggregate level, for the purpose of comparison we will use the state-space methodology and Markov-Switching, as discussed in the previous sections. An advantage of working with disaggregated prices, in these cases, is the ability to better control for them due to changes of weighting in the structures.

5 RESULTS

Evaluation of the models’ performance was conducted using the out-of-sample results for forecasts for up to twelve months forward. All series and models were estimated recursively, starting in January 2008 and ending in March 2012. This means that with every new piece of information added, the models and all their parameters were re-estimated to always obtain the "best" model, conditional on information at that time. In this sense, the sample periods for each estimation varies according to the final sample, i.e. for evaluating the forecasts of the models one step ahead the sample is 51 months, while for forecast twelve months ahead the sample is 40. We preferred to use the full sample for each forecast horizon, because the more informative is the series, the better the results found in MCS.

For the disaggregated models, their subcomponents were re-aggregated using the weightings corresponding to the IPCA estimated for the first month, while for
the remaining months ahead these were estimated based on the result of the forecasts in the previous period.

The loss function used to compare models was the mean squared forecast error (MSFE) which can be defined as the average difference between the estimated value and the actual, squared. The advantage of this measure is that the larger projection deviations receive the highest penalty. It should be reiterated that the MSFE was calculated upon the number index and not in relation to rate of change, so that the results which will be presented later have relatively larger absolute values than if compared in percentage terms.

Initial analysis suggests better MSFE performance on more disaggregated models, while aggregated models showed the worst results. Lower MSFEs were found in the models with the highest level of disaggregation; the model made by forecasting the 52 IPCA items (SARIMA_52) is the one with the lowest MSFE, followed by 9 groups subdivisions (SARIMA_9). The third smallest MSFE was the IPCA was divided into 4 groups: Food and beverages, industrials, services and monitored prices (SARIMA_4). The largest MSFE's were found in the aggregate models in the three forecast methods used (SARIMA, Markov-Switching and structural model). Graphically, the lower performance of the aggregated models compared disaggregated models becomes even clearer (Figure 1), but the model estimated by state-space methodology presents the worst results among all models.

[Figure 1 about here]
We also performed combinations of estimated models, whenever it proved possible that a combination of forecasts resulted in increased accuracy in terms of the loss function. To select the possible combinations, we tested using regressions, following the idea formalized and extended by Chong and Hendry (1986), wherein the effective value is regressed in function of the forecasts obtained on the basis of the estimated models plus a constant, and the estimated coefficients cannot be zero nor one, so that one model overlaps the other. Any value other than that, shows that the joint forecasts provide useful information about the actual value. The weightings for the combination were obtained by regression of the estimates of the values achieved without constant, under the restriction that the sum of the weightings equals one. As shown by Granger and Ramanathan (1984), this method returns the optimal variance-covariance outcome. The weights were estimated for all time horizons in order to always obtain the best combination of templates in each period.

We found five possible combinations that resulted in improved performance of the combined model compared to the results for each of the individual models. Among the aggregate models, the combination of the projections of the model estimated by SARIMA and Markov-switching showed great improvement in performance. Among the disaggregated models, 4 possible combinations were found, but two of them showed no improvement on forecasts for all horizons. The combination of the model disaggregated into 4 groups by the new classification with those disaggregated into 5 or 3 groups showed improvement in all horizons, while the combination of the model disaggregated in 9 groups showed gains at longer horizons, albeit slight. The combination of the model disaggregated in 9 groups also showed improvement when combined with the model disaggregated into 3 groups for the first 3 months and in the
last 5 months. The improvements in the combined results can be better observed in Figure 2.

[Figure 2 about here]

Table 3 below shows the MSFE for all models in forecasts made up to twelve months ahead. The models are ordered from the lowest level of disaggregation to greatest possible. At the end of the table are the results obtained by combining the projections which showed some improvement in terms of the loss function.

[Table 3 about here]

5.1 RESULTS OF MCS

However, only the MSFE analysis alone does not tell us whether the differences in the models are statistically significant, or if the sample period is informative enough to define a better model. Thus, we perform statistical tests to see which models can statistically be considered the best based on the Model Confidence Set selection criteria. The Table 4 shows all p-values resulting from the MCS and those models considered part of the set of "best" models (M*), with 90% and 75% probability, indicated with one and two asterisks respectively, as in Hansen et al (2010).

The results suggest that there is a gain in accuracy in forecasting models with a higher level of disaggregation, because the IPCA projected in the 52 items appears on $\mathcal{M}_{90\%}^*$ in all the forecast horizons. In forecasts of 4, 5, 6, and 7 months
ahead, this model appears as uniquely belonging to $M^*$. Other evidence that also suggests a gain from disaggregating is that the model with the second highest degree of disaggregation, the IPCA projected in all nine groups, appears on $M^*$ in 9 of the 12 forecasts horizons, and for 2, 8, and 12 months forward appears with 90% probability while for 1, 3, 7, 9, 10, and 11 months ahead appears at 75%. The aggregate models (MS, STSM and SARIMA_1) fail to appear in MCS, but the combination of the projections made by the MS and SARIMA_1 models resulted in better performance, making up part of the MCS in projections between 9 and 12 months ahead. The combinations of forecasts from models SARIMA_3 and SARIMA_4N showed an increase in p-value of MCS in relation to models that make up almost all forecast horizons, except for the first three months. The two combinations performed with the model SARIMA_9, the second most disaggregated level, showed p-values of MCS nearly equal to SARIMA_9, except for 1 and 8 months ahead in combination with SARIMA_3.

[Table 4 about here]

It is quite interesting to note that the gains made in the disaggregated models occur at all forecast horizons, both the short term and in the longer spans. It should be noted that the p-values found in the disaggregated models are higher, and the forecast on the highest level of disaggregation appears with a p-value equal to 1 in all 12 forecast horizons.
Remember that all estimated models are univariate and multivariate models that can provide an additional gain. Analysis of multivariate models is one suggestion for future studies on the topic.

5.2 COMPARING MCS WITH TEST OF DIEBOLD & MARIANO (1995)

The methodology of Diebold-Mariano (DM) for comparing models requires a benchmark model for the pair-to-pair comparison of forecasting models. We chose to use as the benchmark the IPCA model composed of 52 items (SARIMA_52), since we find evidence of gains in accuracy in forecasts using the most disaggregated models. Thus, we will compare the mean squared forecast errors for SARIMA_52 against all other models, making sure that the gains are statistically significant for DM. The test proposed by DM has as null hypothesis of equal forecast performance. The test statistic is defined by:

\[ S_1 = \frac{\overline{d}}{\sqrt{\frac{\sum f_d(0)}{T}}} \]  

(11)

where:

- \( d = g(e_{it}) - g(e_{jt}) \) is the differential loss of function;
- \( \overline{d} = \frac{1}{T} \sum_{t=1}^{T} (g(e_{it}) - g(e_{jt})) \);
- \( f_d(0) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_d(r) \), \( \hat{f}_d(0) \) is a consistent estimator \( f_d(0) \);
- \( \gamma_d(r) = E((d_t - \mu)(d_{t-r} - \mu)) \) and \( \sqrt{T} (\overline{d} - \mu) \rightarrow N(0, 2\pi f_d(0)) \).

Thus, the null hypothesis tests at \( E[d_t] = 0 \): that is, the differential function or loss is not significant. The table below presents all tests, using the SARIMA_52 model as a benchmark, and one can observe that the DM test also shows evidence of
performance gain in the forecast for more disaggregated models. In the shorter-term forecasts, to 7 months, only the IPCA group (SARIMA_9) models and combinations thereof are statistically equivalent to at least 10% of significance in relation to the benchmark, with the exception of the 4-month forecast. All other models are inferior to the SARIMA_52 model at a 10% significance level. Over longer terms, there is no clear evidence of improvements from disaggregation, and the only model which declines in forecast performance at all terms is the structural model.

[Table 5 about here]

6 POSSIBLE EXTENSIONS AND LIMITATIONS OF THE WORK

This study examines forecasts from a broader level of disaggregation of Brazilian inflation models, and for comparing the forecasting models we used a recent technique presented by Hansen et al (2010). Nevertheless, there remain some limitations as well as possibilities for various extensions by future studies. One limitation of the study lies in the relatively short sample in historical terms, especially when one considers that the forecasts began in January 2008, losing 51 data points, and also the series experienced some turbulence due to the change of exchange rate regime in January 1999 and the relevant currency devaluation in 2002. Another observation should be made regarding the forecast period, which incorporates several exogenous shocks such as the financial crisis marked by the bankruptcy of Lehman Brothers in September 2008; the sharp rise in commodity prices in 2010, reflecting the
accommodative monetary policies in developed countries; and the European debt crisis, with the first request for the IMF bailout of Greece in April 2010. The projection period also generated a relatively short sample mean square forecast errors, 51 data points for forecasts one month ahead and 40 for data 12 months ahead, which could compromise the evaluation due to the real difference in accuracy of the models studied. However, one of the advantages of MCS is to recognize the limitations of the data, and if these are not adequately explanatory an MCS with all models would be generated. Thus, a possible extension of the work would accomplish the same forecasting exercise for different periods of the IPCA series in order to see whether similar results are obtained in other periods.

Another possible extension to the work would be to incorporate a wider range of forecasting models, including multivariate VAR models, Phillips curve aggregated and disaggregated models, and factorial models, for example. Regarding the model with the highest level of disaggregation, SARIMA_52, it would be possible to estimate each item with the model that best fits the series, not just the SARIMA class of models, which could lead to an even greater gain in disaggregation data.

There is a very large range of possible extensions to the work that could result in additional gains for research methodologies on Brazilian inflation forecasts.

7 CONCLUSION

This paper presents a comparison of inflation forecasts for up to 12 months ahead, as measured by the IPCA, using univariate linear and nonlinear time series, projecting rates from aggregated and disaggregated information. A comparison of the performance of the forecasts was conducted by Model Confidence Set, introduced by Hansen et al (2010), by which it was found that there are significant gains in
disaggregated inflation forecasts, more clearly for shorter forecasting horizons, and less significantly for longer horizons. The analysis by the MSFE and p-value of MCS suggests that disaggregation offers a performance gain even for longer time horizons, although these have not been shown to be statistically significant. One factor that may have limited the results showing significance for longer horizons is the relatively short sample, both in-sample and for the out-of-sample forecasts. The nonlinear models estimated for the aggregated IPCA, Markov-switching, and Structural model showed the worst results among the models. The worst forecast performance was presented by the Structural model, suggesting that models with stochastic coefficients can generate results within the sample that are not repeated out-of-sample. The combination of the forecasts of aggregate models showed significant improvement in the performance of forecasts, something that was not as evident in the combination of the forecasts of disaggregated models.

The results presented in the previous section are somewhat similar to those found in some international work, which showed gains through disaggregation of forecasts, such as Duarte and Rua (2007), who found an inverse relationship between amount of information and forecast horizon: i.e. more disaggregated models had better predictive power in shorter terms. Hubrich and Hendry (2006) also found gains through disaggregation for forecasting inflation in the Eurozone. Sachsida, Ribeiro and Dos Santos (2009) found the Markov-switching model poorly suited to forecast Brazilian quarterly inflation.

The work aims to contribute to the literature emphasizing that the disaggregated analysis of price indices can generate more accurate forecasts. The results are stronger for shorter forecasting horizons.
## APPENDIX I - COMPATIBILIZATION OF STRUCTURES OF IPCA ITEMS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3301</td>
<td>Maintenance and repair services</td>
<td>n.d.</td>
<td>Item 3301 Maintenance and repair services</td>
<td>Item 3301 Maintenance and repair services</td>
</tr>
<tr>
<td>6203</td>
<td>Health insurance</td>
<td>n.d.</td>
<td>Item 6203 Health insurance</td>
<td>Item 6203 Health insurance</td>
</tr>
<tr>
<td>7203</td>
<td>Photography and video recording</td>
<td>Aggregation of sub-items 7201011 Camera and 7201011 Photography accessories</td>
<td>Item 7203 Photography and video recording</td>
<td>Item 7203 Photography and video recording</td>
</tr>
<tr>
<td>7201</td>
<td>Recreation</td>
<td>Item Recreation less sub-items 7201011 Camera and 7201011 Photography accessories</td>
<td>Item 7201 Recreation</td>
<td>Item 7201 Recreation</td>
</tr>
<tr>
<td>8101</td>
<td>Courses</td>
<td>Aggregation of sub-items: 7301004. Textbooks; 7301006.Formal courses; 7301020.Books and technical journals; 7301021.Nursery</td>
<td>Item 8101. Courses less sub-item 8101014. Courses in general</td>
<td>Item 8101 Cursos</td>
</tr>
<tr>
<td>8104</td>
<td>Courses in general</td>
<td>Sub-item 7301007.Courses in general</td>
<td>Sub-item 8101014.Courses in general</td>
<td>Item 8104 Courses in general</td>
</tr>
<tr>
<td>8103</td>
<td>Stationery articles</td>
<td>Aggregation of sub-items: 7301002.Notebooks; 7301003.Stationery articles</td>
<td>Item 8103 Stationery articles</td>
<td>Item 8103 Stationery articles</td>
</tr>
<tr>
<td>9101</td>
<td>Communication</td>
<td>Item 5201 Communication</td>
<td>Item 9101.Communication</td>
<td>Item 9101.Communication</td>
</tr>
</tbody>
</table>

Source: Autor
REFERENCES


Table 1 – IPCA weights of the BCB classifications in changes to the POF

<table>
<thead>
<tr>
<th></th>
<th>Jan-96</th>
<th>Aug-99</th>
<th>Jul-06</th>
<th>Jan-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food and beverages</td>
<td>26.54</td>
<td>22.30</td>
<td>20.29</td>
<td>23.12</td>
</tr>
<tr>
<td>Services</td>
<td>30.43</td>
<td>24.33</td>
<td>23.06</td>
<td>23.66</td>
</tr>
<tr>
<td>Monitored prices</td>
<td>13.25</td>
<td>24.40</td>
<td>31.19</td>
<td>26.70</td>
</tr>
<tr>
<td>Industrials</td>
<td>29.79</td>
<td>28.97</td>
<td>25.46</td>
<td>26.52</td>
</tr>
<tr>
<td><strong>IPCA</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Food at home</td>
<td>18.88</td>
<td>17.15</td>
<td>13.36</td>
<td>15.15</td>
</tr>
<tr>
<td>Monitored prices new</td>
<td>13.22</td>
<td>26.61</td>
<td>29.31</td>
<td>24.51</td>
</tr>
<tr>
<td>Services new</td>
<td>37.53</td>
<td>29.48</td>
<td>31.38</td>
<td>33.72</td>
</tr>
<tr>
<td>Industrials new</td>
<td>30.37</td>
<td>26.76</td>
<td>25.94</td>
<td>26.62</td>
</tr>
<tr>
<td><strong>IPCA</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Services</td>
<td>30.43</td>
<td>24.33</td>
<td>23.06</td>
<td>23.66</td>
</tr>
<tr>
<td>Monitored prices</td>
<td>13.25</td>
<td>24.40</td>
<td>31.19</td>
<td>26.70</td>
</tr>
<tr>
<td>Durable goods</td>
<td>8.25</td>
<td>11.55</td>
<td>10.67</td>
<td>11.56</td>
</tr>
<tr>
<td>Nondurable goods</td>
<td>33.56</td>
<td>31.99</td>
<td>26.04</td>
<td>29.51</td>
</tr>
<tr>
<td>Semidurable goods</td>
<td>14.51</td>
<td>7.72</td>
<td>9.03</td>
<td>8.56</td>
</tr>
<tr>
<td><strong>IPCA</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Tradables</td>
<td>43.92</td>
<td>39.97</td>
<td>33.44</td>
<td>36.13</td>
</tr>
<tr>
<td>Nontradables</td>
<td>42.83</td>
<td>35.63</td>
<td>35.37</td>
<td>37.16</td>
</tr>
<tr>
<td>Monitored prices</td>
<td>13.25</td>
<td>24.40</td>
<td>31.19</td>
<td>26.70</td>
</tr>
<tr>
<td><strong>IPCA</strong></td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Source: IBGE; Banco Central do Brasil

Table 2 – Estimated models and their breakdowns

<table>
<thead>
<tr>
<th>Name</th>
<th>Composition</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARIMA_1</td>
<td>Overall Index</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_9</td>
<td>9 groups</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_52</td>
<td>52 items</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_4</td>
<td>Industrials old, Services old, Monitored old and group Food and Beverages</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_3</td>
<td>Tradables, Non-tradables, Monitored old</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_5</td>
<td>Durables, Semidurables, Non-durables, Services old and Monitored old</td>
<td>SARIMA</td>
</tr>
<tr>
<td>SARIMA_4N</td>
<td>Industrials new, Services new, Monitored new and subgroup Food at home</td>
<td>SARIMA</td>
</tr>
<tr>
<td>MS</td>
<td>Overall Index</td>
<td>Markov-Switching</td>
</tr>
<tr>
<td>STSM</td>
<td>Overall Index</td>
<td>Structural Model</td>
</tr>
</tbody>
</table>

Source: Autor
Table 3 - Mean Square Forecast Error of the models for the forecast horizon

<table>
<thead>
<tr>
<th>Model</th>
<th>Forecast Horizon (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>MS</td>
<td>0.219</td>
</tr>
<tr>
<td>STSM</td>
<td>0.216</td>
</tr>
<tr>
<td>SARIMA_1</td>
<td>0.146</td>
</tr>
<tr>
<td>SARIMA_3</td>
<td>0.140</td>
</tr>
<tr>
<td>SARIMA_4</td>
<td>0.134</td>
</tr>
<tr>
<td>SARIMA_4N</td>
<td>0.143</td>
</tr>
<tr>
<td>SARIMA_5</td>
<td>0.143</td>
</tr>
<tr>
<td>SARIMA_9</td>
<td>0.128</td>
</tr>
<tr>
<td>SARIMA_52</td>
<td>0.098</td>
</tr>
<tr>
<td>SARIMA_1 x MS</td>
<td>0.652</td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_3</td>
<td>0.134</td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_5</td>
<td>0.136</td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_9</td>
<td>0.128</td>
</tr>
<tr>
<td>SARIMA_3 x SARIMA_9</td>
<td>0.121</td>
</tr>
</tbody>
</table>

Note: The shaded areas and in bold denotes de minimum EQM

Table 4 - Model Confidence Set and p-value of the estimated models

<table>
<thead>
<tr>
<th>Sample (n)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>0.018</td>
<td>0.016</td>
<td>0.011</td>
<td>0.002</td>
<td>0.002</td>
<td>0.009</td>
<td>0.022</td>
<td>0.052</td>
<td>0.059</td>
<td>0.055</td>
<td>0.063</td>
<td>0.049</td>
</tr>
<tr>
<td>STSM</td>
<td>0.005</td>
<td>0.009</td>
<td>0.006</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
<td>0.006</td>
<td>0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>SARIMA_1</td>
<td>0.059</td>
<td>0.042</td>
<td>0.019</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.013</td>
<td>0.031</td>
<td>0.025</td>
<td>0.006</td>
<td>0.005</td>
</tr>
<tr>
<td>SARIMA_3</td>
<td>0.106 **</td>
<td>0.122 **</td>
<td>0.030</td>
<td>0.015</td>
<td>0.011</td>
<td>0.027</td>
<td>0.051</td>
<td>0.127 **</td>
<td>0.142 **</td>
<td>0.164 **</td>
<td>0.127 **</td>
<td>0.305 *</td>
</tr>
<tr>
<td>SARIMA_4</td>
<td>0.133 **</td>
<td>0.206 *</td>
<td>0.152 **</td>
<td>0.049</td>
<td>0.032</td>
<td>0.038</td>
<td>0.070</td>
<td>0.190 **</td>
<td>0.171 **</td>
<td>0.193 **</td>
<td>0.120 **</td>
<td>0.240 *</td>
</tr>
<tr>
<td>SARIMA_4N</td>
<td>0.024</td>
<td>0.242 *</td>
<td>0.152 **</td>
<td>0.049</td>
<td>0.026</td>
<td>0.023</td>
<td>0.063</td>
<td>0.093</td>
<td>0.106 **</td>
<td>0.154 **</td>
<td>0.152 **</td>
<td>0.431 *</td>
</tr>
<tr>
<td>SARIMA_5</td>
<td>0.059</td>
<td>0.035</td>
<td>0.038</td>
<td>0.005</td>
<td>0.004</td>
<td>0.012</td>
<td>0.031</td>
<td>0.103 **</td>
<td>0.118 **</td>
<td>0.140 **</td>
<td>0.120 **</td>
<td>0.431 *</td>
</tr>
<tr>
<td>SARIMA_9</td>
<td>0.140 **</td>
<td>0.262 *</td>
<td>0.183 **</td>
<td>0.062</td>
<td>0.066</td>
<td>0.095</td>
<td>0.151 **</td>
<td>0.286 *</td>
<td>0.187 **</td>
<td>0.199 **</td>
<td>0.183 **</td>
<td>0.521 *</td>
</tr>
<tr>
<td>SARIMA_52</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
<td>1.000 *</td>
</tr>
<tr>
<td>SARIMA_1 x MS</td>
<td>0.068</td>
<td>0.067</td>
<td>0.006</td>
<td>0.006</td>
<td>0.014</td>
<td>0.037</td>
<td>0.078</td>
<td>0.106 **</td>
<td>0.128 **</td>
<td>0.112 **</td>
<td>0.254 *</td>
<td></td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_3</td>
<td>0.097</td>
<td>0.198 **</td>
<td>0.152 **</td>
<td>0.062</td>
<td>0.041</td>
<td>0.049</td>
<td>0.087</td>
<td>0.198 **</td>
<td>0.180 **</td>
<td>0.199 **</td>
<td>0.183 **</td>
<td>0.464 *</td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_5</td>
<td>0.080</td>
<td>0.198 **</td>
<td>0.152 **</td>
<td>0.062</td>
<td>0.032</td>
<td>0.038</td>
<td>0.070</td>
<td>0.127 **</td>
<td>0.160 **</td>
<td>0.193 **</td>
<td>0.178 **</td>
<td>0.464 *</td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_9</td>
<td>0.140 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.199 **</td>
<td>0.183 **</td>
<td>0.521 *</td>
</tr>
<tr>
<td>SARIMA_3 x SARIMA_9</td>
<td>0.148 **</td>
<td>0.262 *</td>
<td>0.183 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.301 *</td>
<td>0.196 **</td>
<td>0.199 **</td>
<td>0.183 **</td>
</tr>
</tbody>
</table>

Note: The forecasts in $\hat{\sigma}_p^2$ is identified by one asterisk and p-value in bold in the shaded area, while those in $\hat{\sigma}_{tsu}^2$ is identified by two asterisks and p-value in the shaded area.

Table 5 - Diebold-Mariano test of model more disaggregated than others, by time horizon
### Table: Mean Square Forecast Errors of the estimated models by forecast horizon

<table>
<thead>
<tr>
<th>Forecast Horizon (months)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample (n)</td>
<td>51</td>
<td>50</td>
<td>49</td>
<td>48</td>
<td>47</td>
<td>46</td>
<td>45</td>
<td>44</td>
<td>43</td>
<td>42</td>
<td>41</td>
<td>40</td>
</tr>
<tr>
<td>MS</td>
<td>-0.12**,-0.37**, -0.70**,-1.11**,-1.47**,-1.80**,-2.09**, -2.52**,-3.20**,-4.11**,-4.69**,-5.20**</td>
<td>(0.0526),(0.1461),(0.2646),(0.4281),(0.5657),(0.9244),(1.2160), (1.5488),(2.0148),(2.6423),(3.1509),(3.7217)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STSM</td>
<td>-0.12***,-0.57***,-1.37***,-2.42***,-3.63***,-4.90***,-5.81***,-7.03***,-8.70***,-10.40***,-11.26***,-11.36**</td>
<td>(0.0303),(0.1476),(0.3571),(0.5652),(0.7663),(1.1464),(1.6931),(2.3134),(3.0126),(3.7005),(4.1362),(4.4567)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_1</td>
<td>-0.05**,-0.25**,-0.65**,-1.22**,-1.87**,-2.52**,-3.07**,-3.55**,-3.92**,-4.53**,-5.36**,-6.33**</td>
<td>(0.0213),(0.1000),(0.2567),(0.4576),(0.7327),(1.1277),(1.6273),(2.1543),(2.6388),(3.2988),(3.9243),(4.3570)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_3</td>
<td>-0.04**,-0.20**,-0.49**,-0.82**,-1.02**,-1.18**,-1.39**,-1.41**,-1.53**,-1.55**,-1.63**,-1.76**</td>
<td>(0.0201),(0.0910),(0.2199),(0.3396),(0.4448),(0.5676),(0.7679),(0.9645),(1.1568),(1.2641),(1.2979),(1.3427)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_4</td>
<td>-0.04**,-0.13**,-0.24**,-0.53**,-0.54**,-0.63**,-0.70**,-0.71**,-0.86**,-1.03**,-1.36**,-1.47**</td>
<td>(0.0177),(0.0652),(0.1224),(0.1991),(0.2424),(0.3228),(0.4122),(0.4724),(0.5462),(0.6461),(0.7468),(0.8752)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_4N</td>
<td>-0.04**,-0.14*,-0.32**,-0.59**,-0.75**,-0.89**,-0.99**,-1.15**,-1.34**,-1.40**,-1.43**,-1.13**</td>
<td>(0.0221),(0.0654),(0.1643),(0.2390),(0.2838),(0.3721),(0.4824),(0.5948),(0.7519),(0.9221),(1.0803),(1.0619)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_5</td>
<td>-0.04**,-0.22***,-0.49**,-0.86***,-1.04***,-1.25***,-1.29**,-1.22**,-1.33**,-1.40**,-1.55**,-1.54**</td>
<td>(0.0169),(0.0815),(0.1962),(0.3095),(0.3752),(0.4948),(0.5362),(0.6911),(0.8620),(1.1041),(1.4048)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_9</td>
<td>-0.03,-0.07,-0.18,-0.32**,-0.39**,-0.41,-0.41,-0.45,-0.56,-0.67,-0.71**,-0.57**</td>
<td>(0.0167),(0.0620),(0.1231),(0.1789),(0.2494),(0.2931),(0.3457),(0.3919),(0.4362),(0.5362),(0.6714),(0.7009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_1 x MS</td>
<td>-0.24**,-0.52**,-0.83**,-1.12**,-1.43**, -1.58,-1.68,-1.70,-1.92,-1.98,-2.03**</td>
<td>(0.0099),(0.0232),(0.3259),(0.4777),(0.7510),(0.9572),(1.0599),(1.0771),(1.2151),(1.1827),(1.3094)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_3</td>
<td>-0.04**,-0.14*,-0.27**,-0.48**,-0.59**,-0.71**,-0.80,-0.85,-0.96**,-1.00,-1.04,-0.94**</td>
<td>(0.0164),(0.0760),(0.1581),(0.2543),(0.3113),(0.4048),(0.5363),(0.7169),(0.8676),(0.9889),(0.9622)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_5</td>
<td>-0.04**,-0.14*,-0.28**,-0.50**,-0.63**,-0.77**,-0.86**,-0.93**,-1.08**, -1.16,-1.21,-0.97**</td>
<td>(0.0169),(0.0766),(0.1575),(0.2707),(0.3141),(0.3953),(0.4574),(0.5031),(0.6237),(0.7948),(0.9742),(1.0964)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_4N x SARIMA_9</td>
<td>-0.03,-0.07,-0.15,-0.36,-0.49,-0.64,-0.69,-0.53</td>
<td>(0.0188)</td>
<td>(0.0584)</td>
<td>(0.1259)</td>
<td>(0.4568)</td>
<td>(0.5118)</td>
<td>(0.5329)</td>
<td>(0.6321)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARIMA_3 x SARIMA_9</td>
<td>-0.02,-0.07,-0.15,-0.36,-0.49,-0.64,-0.69,-0.53</td>
<td>(0.0165),(0.0586),(0.1259),(0.4568),(0.5118),(0.5329),(0.6321)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The coefficients are the constant from the regression of the difference between the MSE of the most disaggregated model (SARIMA_52) and all other models with a constant. The *** stands for a significance level of 1%, ** stands for a significance level of 5% and * stands for a significance level of 10%.

**Figure 1** – Mean Square Forecast Errors of the estimated models by forecast horizon
Figure 2 – Mean Square Forecast Errors of the combined models estimated by forecast horizon