

APRIMORANDO A PREVISÃO DE PREÇOS DE ALUMÍNIO POR MEIO DE COMBINAÇÕES DE PREVISÕES, MODEL CONFIDENCE SET E SARIMAX

*IMPROVING ALUMINUM PRICE PREDICTABILITY FROM FORECAST COMBINATIONS, MODEL
CONFIDENCE SETS AND SARIMAX AND FORECAST COMBINATION*

João Bosco Barroso de CASTRO

Universidade de São Paulo
joabosco.castro@usp.br

Alessandra de Ávila MONTINI

Universidade de São Paulo
amontini@usp.br

Emerson Fernandes MARÇAL

Fundação Getúlio Vargas de São Paulo (FGV-SP)
emerson.marc@fgv.br

Recebido em 05/2019 – Aprovado em 06/2020

Resumo

Metais industriais levaram o preço das commodities a um patamar mais elevado em 2017, com alta de 28%. Entre os metais industriais, alumínio destaca-se por ser o segundo metal mais usado após o aço. Como o alumínio é um insumo fundamental para uma variedade de aplicações industriais, grandes oscilações nos preços podem ter forte impacto nos termos de troca. Melhorar a acuracidade das previsões de preços do alumínio é crítico para agentes de políticas econômicas, produtores, consumidores industriais e investidores. Este trabalho propõe uma metodologia inédita para obtenção de uma combinação ótima de modelos de previsão de preços de alumínio utilizando Combinações de Previsões, Model Confidence Set e Sarimax. Os resultados mostraram que o melhor modelo foi a combinação com Sarimax e Sarima. Conclui-se também que estoques e preços de alumínio de 3 meses são variáveis-chave para melhorar a acuracidade das previsões.

Palavras-chave: preços de commodities, combinação de previsões, model confidence set

Abstract

Industrial metals led the commodity space higher in 2017, closing the year up 28%. Among industrial metals, aluminum is the second most used metal in the world after steel. As aluminum is a key input for a variety of industrial applications, large swings in prices can have a major impact on the terms of trade. Therefore, improve forecasts on aluminum prices is of critical importance to economic policy decision makers, producers, industry consumers and investors. This work proposes a novel approach to obtain an optimal combination for aluminum price forecasting by utilizing Forecast Combination, Model Confidence Set and Sarimax. Five individual models were estimated. The results showed that the best model was the forecast combination including Sarimax and Sarima models. The work indicated that aluminum inventories and 3-month aluminum forward prices improved price forecast accuracy, as key covariates to improve policy decisions.

Keywords: *Commodity prices; forecast combination; model confidence set*

INTRODUÇÃO

Commodity markets benefited from strong economic activity, supply disruptions, and increased geopolitical risk in 2017, helping to push commodity investor to multi-year highs. Industrial metals, such as aluminum and zinc, led the commodity space higher in 2017, closing the year up 28%, while precious metals increased 10%. The laggards during 2017 included soft commodities (-16%), such as edible oil, and agriculture (-12%).

Among industrial metals, aluminum stands out as it is the second most used metal after steel for

modern societies. It has the widest diversity of end-use applications compared to any other metal. Semi-finished steel and refined copper, for example, are both heavily reliant on the construction sector. The diversity of aluminum applications lends itself to a stronger demand performance once investment demand falls, especially in emerging markets.

Recent years have witnessed a significant increase of aluminum use in transportation, building, packaging, and electrical engineering due to its versatile properties. Life cycle results show that, in automotive applications, each kilogram of aluminum replacing mild steel, cast iron or high strength steel saves between 13 to 20 kg of greenhouse gas emissions (Bertram, 2009). In 2010, about 30% of aluminum used globally was used in transportation. Given its infinite recycling capacity, 75% of all aluminum ever produced is currently still in productive use.

Aluminum is the most heavily traded nonferrous metal on the London Metals Exchange (LME), that sets its spot and futures prices. The LME is the largest pure commodity exchange in Europe and the world's tenth largest futures exchange. The spot price is highly volatile, with the standard deviation of annual differences in the logarithm of the price equal to approximately 0.28 over the last three decades (Baldursson, 1999).

Several studies were developed for commodity price forecasting, including approaches for projecting both the level as well as structural price breaks (Leuthold et al., 1970; Driehuis, 1976). Such approaches were carried out by the World Bank, as described by Duncan (1984) and Labys and Pollack (1984). Kouassi, Labys and Colyer (1998) showed how univariate models can be applied to commodities. Labys and Kouassi (2004) developed structural models for short and medium-term forecasts by identifying cycles and structural breaks. Several authors produced

works using time series techniques for forecasting cycles, like Harvey and Jaeger (1993) and Mills (2003).

Since the pioneering work of Bates and Granger (1969) on forecast combinations, different studies have shown that combining forecasts provide results that are superior to those of individual models, including those by Newbold and Granger (1974), Makridakis and Winkler (1983), Hendry and Clements (2002), Stock and Watson (2001, 2004), Pesaran and Timmermann (2005) and Issler and Lima (2009).

This work seeks to improve aluminum price forecasting by utilizing the Model Confidence Set (Hansen, Lunde and Nason, 2011) for selecting the set of dominant forecast combination models over individual models. Individual model estimations were based on different SARIMAX and structural models. For SARIMAX estimations, 30 independent variables and 37 transformations for each variable were assessed, including weak exogeneity and Granger causality tests, leading to four selected co-variables, namely Brent crude oil price, non-ferrous metals index, 3-month aluminum forward price and world aluminum inventories. The sample used for adjusting the parameters of each individual model was a 36-month rolling window.

Out-of-samples forecasts were based on 12-month ahead time horizon for each rolling window. Forecast combinations were performed using seven approaches to determine weights for individual models, namely (i) simple average; (ii) ordinary least squares; (iii) weighted mean, based on the inverse of the root mean square error; (iv) weighted mean based on the Akaike information criterion; (v) minimization of the sum of the squares of the residuals, with the restriction that the sum of the weights is 1; (vi) minimization of the sum of the sum of the squares of the residuals, with the restriction that the sum of the

weights is 1 and each weight is not negative; (vii) minimization of the sum of the sum of the squares of the residuals, with the restriction that the sum of the weights is not negative.

The results showed that the best model was the forecast combination including SARIMAX and SARIMA models, adjusted by ordinary least squares. The work also indicated that world aluminum inventories and 3-month aluminum forward prices, as explanatory variables in the SARIMAX model, improved aluminum price forecast accuracy. Inventories and forward prices were more relevant to improve model predictability for the latest 36-month out-of-sample forecasts, compared to 48 and 60-month out-of-sample forecasts.

This paper is structured as follows: after this introductory section, the second topic deals with the literature review, including forecast combination and the Model Confidence Set. The third section presents data treatment and independent variables for SARIMAX models. The fourth section provides the results, including individual models, forecast combinations and comparisons of the individual and combined models. The fifth section concludes by focusing on work's contributions, limitations and suggesting future studies.

FORECAST COMBINATION

LITERATURE REVIEW

The main argument favoring combining forecasts are: (i) the true data generating process is unknown. Even the most complicated model is likely to be misspecified and can, at best, provide a reasonable local approximation. It is highly unlikely that a single model will dominate uniformly over time; (ii) the best model may change over time in a way that can be difficult to

track on the basis of past forecasting performance. Combining forecasts across different models may be viewed as a way to make the forecast more robust against misspecification biases and measurement errors in the data set; (iii) it is possible that diversification gains from combining across a set of forecasting models will dominate the strategy of only using a single forecasting model.

The first theoretical approach for combining forecasts was developed by Bates and Granger (1969), based on the concept of the diversification gains of a portfolio. The model assumes that the error variances of two forecasts, σ_1 and σ_2 , are constant over time and that neither forecast is biased. The problem lies in determining the weights of the individual forecasts that generate a combined model that minimizes combination variance.

The concept used by Bates and Granger (1969) was reinforced by Nelson (1972), through a relatively independent development, with application in the North American economy. Newbold and Granger (1974) maintained all the suppositions and combination methods of Bates and Granger (1969), and increased the number of individual forecasts, based on ARIMA, Holt-Winters and stepwise autoregression models. Newbold and Granger (1974) concluded that the predictive performance of the combination of the three models was superior to that of the individual models. Makridakis and Winkler (1983) analyzed the accuracy of forecast combinations using 5 weighted mean methods, based on 10 individual models. The two methods that were most accurate were weighting by way of the inverse of the sum of squared errors. The second was based on the exponential smoothing of the weights obtained in the first method. The authors emphasized that the accuracy of the combination depends on how different the individual models

are, but the performance of the combined projection reaches saturation point with four or five individual forecasts.

A different method was proposed by Engle, Granger and Kraft (1984). In this work, the traditional weight methods were used in the combinations of two inflation forecasts, based on error variances and covariances, but the variances developed were modeled conditional upon past values, using the ARCH method. Despite being potentially useful, this method proved to be more complex and less attractive than the usual combination techniques for minimizing errors.

Granger and Ramanathan (1984) stressed that conventional models of linear forecast combination could be viewed as a structured form of regression. They concluded that the linear combination method is equivalent to the ordinary least squares method, assuming the combined forecast as a dependent variable and the individual forecasts as independent variables.

Granger and Newbold (1986) argued that point forecasts should be combined in a linear way, using some mechanism for estimating the weights, which are not necessarily positive or normalized (in other words, the sum of the weights is 1), with the objective of minimizing the variance of the residuals.

Christoffersen and Diebold (1997) investigated the case in which the bias is proportional to the conditional variance of the dependent variable. Assuming the forecasts to be similarly volatile, the authors argued that, if a loss function is more symmetrical than those used in constructing individual forecasts, it is possible to obtain a combination with a greater predictive performance than individual forecasts.

Stock and Watson (2001) found evidence for the use of forecast combinations, particularly by the mean and median and weighted by the inverse of

the mean squared error, with dominance of the combination for forecast time horizons of one, six and twelve months. Hendry and Clements (2002) argued that forecast combinations can produce favorable empirical results because they provide a type of insurance against externalities and structural breaks.

Subsequent works dealt with various topics, like those of Chan et al. (2004), on the use of variable weights; Stock and Watson (2004), on the fact that complex combinations are not always more accurate than combinations using the average; Elliott and Timmermann (2005), relative to the use of regime change models; Liang, Lee and Shao (2006), with regard to linear combination models; Fan and Deng (2007), on using forecast errors for constructing variable weights; Kim (2008), on generalized autoregression models; and Smith & Wallis (2009), on the error of finite samples in estimating the combined weight.

Issler and Lima (2009) proposed a new approach for forecast combinations, using bias correction, which proved potentially superior to other techniques in an asymptotic context. Kolassa (2011) emphasized the construction of weights, based on the Akaike information criterion (AIC). Bruggeman and Luktepohl (2013) showed that the combined forecast, based on minimizing the mean squared error, was more accurate when using aggregate GDP and monetary base data of the United States than when using disaggregated data. Hsiao and Wan (2014) compared combinations using AIC weighting, regression, the Bayesian method and the minimization of squared errors. Recent topics and topics for future research deal with aspects such as probability density combinations, varying weights over time and combinations with asymmetric loss functions.

FORECAST COMBINATION MODEL

Given a time series y_t and a set of M forecasting models denoted $F_i()$, $i = 1, \dots, M$, with possible inputs, the following relationship is presumed for each model i :

$$y_t = F_i(Y_{t-1}) + \varepsilon_{i,t}$$

where Y_{t-1} denotes a vector of time series observations up to time $t-1$ and $\varepsilon_{i,t}$ are IID disturbances with respect to t , assumed $E[\varepsilon_{i,t}] = 0$. Independently, each model $F_i()$ is fit to produce a fitted forecast model that is denoted by $\hat{F}_i()$. Consider the following notation for the results of $\hat{F}_i()$:

$\hat{y}_{i,t}$ denotes the forecast for fitted model $\hat{F}_i()$ at time t ;

$\hat{e}_{i,t} = y_t - \hat{y}_{i,t}$ denotes the prediction error for fitted model $\hat{F}_i()$ at time t

$\hat{\sigma}_{i,t}^2 = Var(\hat{e}_{i,t})$ denotes the prediction error variance for fitted model $\hat{F}_i()$ at time t

The combined forecast at time t , denoted by $\hat{y}_{c,t}$, uses combination weights, \hat{w}_i , associated with predictions for each fitted model $\hat{F}_i()$ at time t , denoted by $\hat{y}_{i,t}$, as follows:

$$\hat{y}_{c,t} = \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t}$$

The combined forecast prediction error and variance are generally defined by:

$$\hat{e}_{c,t} = y_t - \hat{y}_{c,t}$$

$$Var(\hat{e}_{c,t}) = Var(y_t - \hat{y}_{c,t}) = Var(y_t(1 - \sum_{i=1}^M \hat{w}_i) + \sum_{i=1}^M \hat{w}_i \hat{e}_{i,t})$$

When we assume or impose that the constraint $\sum_{i=1}^M \hat{w}_i = 1$, simplification results in:

$$Var(\hat{e}_{c,t}) = \sum_{i=1}^M \hat{w}_i \hat{e}_{i,t}$$

$$= \hat{W}_t' \Sigma \hat{W}_t$$

where \hat{W}_t denotes the fitted weight vector $M \times 1$ and $\Sigma = Var(E_t)$, with $E_t = (\hat{e}_{1,t}, \hat{e}_{2,t}, \dots, \hat{e}_{M,t})'$ denotes the ensemble prediction error vector at time t .

Alternatively, this can be expressed as:

$$Var(\hat{e}_{c,t}) = \sum_{i=1}^M \hat{w}_i^2 Var(\hat{e}_{i,t})$$

$$+ 2 \sum_{j < i}^M \hat{w}_i \hat{w}_j Cov(\hat{e}_{i,t}, \hat{e}_{j,t})$$

$$= \sum_{i=1}^M \hat{w}_i^2 \hat{\sigma}_{i,t}^2 + 2 \sum_{j < i}^M \hat{w}_i \hat{w}_j \hat{\rho}_{i,j,t} \sigma_{i,t} \sigma_{j,t}$$

If the prediction errors for the individual forecasts are not correlated at time t , then:

$$\Sigma = diag(\hat{\sigma}_{i,t}^2)$$

and the combined forecast prediction error variance and standard errors are:

$$Var(\hat{e}_{c,t}) = \sum_{i=1}^M \hat{w}_i^2 \hat{\sigma}_{i,t}^2$$

$$\sigma_{c,t} = \sqrt{\sum_{i=1}^M \hat{w}_i^2 \hat{\sigma}_{i,t}^2}$$

Forecast weights $\hat{w}_i, i = 1, \dots, M$, of M individual models can be determined using the following methods:

Simple combinations, in three categories, as per Clark and McCracken (2006): simple mean; median; trimmed mean.

Weights defined by the root mean squared error;

Weights corrected by the Akaike information criterion;

Weights determined by the ordinary least squares method;

Weights determined by the restricted least squares criterion, that is, non-negative least squares; equally restricted least squares; equally restricted and non-negative least squares.

Simple Combinations

Assuming simple mean, the estimated weights for individual forecasts are given by:

$$\hat{w}_i = \frac{1}{M}, \forall i \in \hat{F}_i()$$

The combined forecast can also utilize the sample median as the estimated weight:

$$\hat{y}_{c,t} = \text{median} \{ \hat{F}_i() \}$$

$$\hat{w}_i = \frac{1/\text{RMSE}_i}{W}, \quad \text{where } W = \sum_{i=1}^M \frac{1}{\text{RMSE}_i}$$

Using the trimmed mean, the weights of the maximum and minimum forecasts are equalized to zero. For the other models, the weight:

$$\hat{w}_i = \frac{1}{M-2}$$

is attributed, and the combined forecast is given by:

$$\hat{y}_{c,t} = \frac{1}{M-2} \sum_{i=1}^{M-2} \hat{y}_{i,t}$$

Among three simple combination methods above, the simple mean is the most representative measure of the central trend if the forecasts belong to the same probability distribution; if not, the median or trimmed mean can supply the most suitable combined forecast.

The Root Mean Squared Error Method

This method uses the inverse of the mean squared error (MSE) of the series of forecast errors of each model, in relation to the sum of the inverse of the RMSE of all the models, as the weight of each individual forecast. Therefore, the smaller the RMSE, the smaller the weight associated with forecast $\hat{y}_{i,t}$ of the $F_i()$ model, in time t , where $i = \{1, \dots, M\}$.

Therefore, the estimated weight for forecast $\hat{y}_{i,t}$ of model $F_i()$ will be:

Akaike Information Criterion Method (AIC)

This method uses the Akaike Information Criterion (AIC) associated with the series of forecast errors of each model as the weight of each individual forecast.

Let the AIC of model $F_i()$ be given by: $AIC_i = AIC(\hat{e}_{i,t-j}; P_i)$, $\hat{e}_{i,t-j}$ is the forecast error series of the adjusted model $\hat{F}_i()$ at t , for h steps ahead, $j = \{h-1, \dots, 0\}$ and P_i is the number of parameters of the model. Assuming $h=1$, $AIC_i = AIC(\hat{e}_{i,t}; P_i)$.

Let the smaller AIC_i be denoted by AIC_{min} and it is defined as:

$$\Delta_i = AIC_i - AIC_{min}$$

The estimated weight for the forecast $\hat{y}_{i,t}$ of model $F_i()$ will be:

$$\hat{w}_i = \frac{\exp(-\frac{1}{2} \lambda \Delta_i)}{W},$$

$$\text{Where } W = M \sum_{i=1}^M \left[\exp(-\frac{1}{2} \lambda \Delta_i) \right] \text{ e } 0 \leq \lambda$$

When $0 \leq \lambda \leq 1$, the resulting weight reflects a value between the weight obtained by the simple mean (when $\lambda = 0$) and the usual weight by the AIC (when $\lambda = 1$). When λ is raised to 10, for example, the method tends to attribute a greater weight to the model with the smaller AIC. The

trend of selecting the best model is frequently observed when $\lambda = 1$, unless other forecasts have AIC that are relatively close to AICmin. The combined forecast at t will be given by:

$$\hat{y}_{c,t} = \sum_{i=1}^M \frac{\exp(-\frac{1}{2}\lambda\Delta_i)}{W} \hat{y}_{i,t}$$

Ordinary Least Squares Method (OLS)

The weights are estimated by way of the OLS regression of T observed values of the dependent variable y_t on M individual forecasts for each instant j of estimation period T, where $j = \{1, 2, \dots, T\}$. In the case of forecasting h steps ahead, we have $t = T + h$, where t is the instant of time of the last forecast. The general expression of OLS is given by:

$$y_t = w_0 + w_1\hat{y}_{1,t} + w_2\hat{y}_{2,t} + \dots + w_M\hat{y}_{M,t} + \epsilon_t, \text{ where } t = \{1, 2, \dots, T\}$$

The estimated weights \hat{w}_i are given by:

$$\hat{w}_i = \sum_{t=1}^T (\hat{y}'_t \hat{y}_t)^{-1} \sum_{t=1}^T (\hat{y}'_t y_t), \text{ where } i = \{1, \dots, M\}$$

The combined forecast at t will be given by:

$$\hat{y}_{c,t} = \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t}$$

According to Granger & Ramanathan (1984), three versions of the OLS method can be considered:

$$y_t = \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t} + \epsilon_{i,t}, \text{ t.q. } \sum_{i=1}^M \hat{w}_i = 1$$

$$y_t = c + \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t} + \epsilon_{i,t}, \text{ t.q. } \sum_{i=1}^M \hat{w}_i = 1$$

$$y_t = c + \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t} + \epsilon_{i,t}$$

The first equation is a restricted regression, where the sum of the weights must add the unit and the individual forecasts have no bias in order to ensure that the combination of forecasts is also not biased. The second equation is a variation of the first, where the individual forecasts can have a bias, which is corrected by the introduction of the constant c. The third equation is a generalization of the two previous ones, where the sum of the weights is not necessarily the unit. The third general version of the OLS method was one of the methods used for the combined forecast of aluminum prices.

Restricted Least Squares Method

The three methods of this category are formulated as problems of least squares subject to restrictions and solved by way of the optimization process. The three methods minimize the objective function:

$$z = \sum_{t=1}^T (y_t - \sum_{i=1}^M \hat{w}_i \hat{y}_{i,t})^2$$

y_t is the observed value;

\hat{w}_i is the weight for model i;

$\hat{y}_{i,t}$ is the estimated value of the model at instant t.

Depending on the restrictions, there are three methods:

Non-negative least squares, which impose the restriction: $\hat{w}_i \geq 0, i = 1, \dots, M$

Equally restricted least squares, which impose the restriction:

$$\sum_{i=1}^M \hat{w}_i = 1, i = 1, \dots, M$$

Equally restricted and non-negative least squares, which impose the restriction:

$$\sum_{i=1}^M \hat{w}_i = 1, \hat{w}_i \geq 0, i = 1, \dots, M$$

INDIVIDUAL MODEL ESTIMATION

Three approaches were used for estimating individual models to arrive at forecast combinations, namely: SARIMA, SARIMAX and structural modeling. Since SARIMA can be viewed as a particular case of SARIMAX, only the later was included in this section.

SARIMAX

SARIMAX can be expressed by a SARIMA model with exogenous variables, according to (2.2), in which X_t represents the stationary series, including an autoregressive seasonal component $\sum_{j=1}^P \Phi_j X_{t-j}$ of order P and seasonal period s and a moving average seasonal component $\sum_{k=1}^Q \theta_k Z_{t-k}$ of order Q and seasonal period s. According to Stoffer and Shumway (2006), the general form of SARIMAX is expressed by:

$$X_t = \Gamma U_t + \sum_{j=1}^P \phi_j X_{t-j} + \sum_{j=1}^P \Phi_j X_{t-j} + \sum_{k=1}^q \theta_k Z_{t-k} + \sum_{k=1}^Q \Theta_k Z_{t-k} + \gamma_{2t} - \gamma_{3t} + \varphi_t, \varphi_t \sim NI(0, \sigma_\varphi^2)$$

where:

Γ : is the k x r vector of the coefficients of the exogenous variables;

U_t : is the r x 1 vector of the exogenous variables for representing current and past values.

3.2 Structural model

Formally proposed by Harvey (1989), the structural model models a univariate series as a function of its non-observable components: trend; seasonality; and cycle following a linear approach. The structural denomination arises from the fact of looking for a model by way of components that clearly interpret the reality. The generalized additive linear model is expressed by:

$$X_t = \mu_t + \gamma_t + \omega_t + Z_t,$$

where $\mu_t, \gamma_t, \omega_t$ are the stochastic components of trend, seasonality and cycle, and Z_t is white noise, as described in Section 2.2.1. The $\mu_t, \gamma_t, \omega_t$ components follow the formation laws shown below. The parameters to be estimated are the $\sigma_v^2, \sigma_\eta^2, \sigma_\phi^2, \sigma_\psi^2, \sigma_\xi^2$ variances. The seasonal component generally requires (s-1) state equations, where s is the seasonal period; in the case below, it was considered that s = 4.

Trend

$$\mu_t = \mu_{t-1} + \beta_t + v_t, v_t \sim NI(0, \sigma_v^2)$$

$$\beta_t = \beta_{t-1} + \eta_t, \eta_t \sim NI(0, \sigma_\eta^2)$$

Seasonality

$$\gamma_{2t} = \gamma_{2t-1}$$

$$Y_{3t} = Y_{3t-1}$$

Cycle

$$\omega_t = \lambda_t \psi_t, \psi_t \sim NI(0, \sigma_\psi^2)$$

$$\psi_t = \psi_{t-1} + \xi_t, \xi_t \sim NI(0, \sigma_\xi^2)$$

THE MODEL CONFIDENCE SET

Unlike conventional models for model selection, the Model Confidence Set (MCS), as proposed by Hansen, Lunde & Nason (2011), can make multiple comparisons of models, based on a given descriptive level. This fact enables a set of models to be selected that gives comparable forecasting accuracy at a given descriptive level, as defined by the MCS.

The objective of the Model Confidence Set (MCS) is to determine the set of models M^* , which contains the best model(s) from a collection of models M^0 , in which the concept of “best” is defined on the basis of a criterion that is associated with the forecast performance. MCS estimates a set \widehat{M}^* , which is the set that contains the best models for a given descriptive level.

An important contribution of MCS is that this approach recognizes data limitations. Data with the same information capacity result in an \widehat{M}^* that contains just one model. Data with less information capacity result in a set of models with a forecast performance that is comparable to a given descriptive level of the MCS. Therefore, the MCS differs from conventional model selection methods, which determine just one model as being the best among all the adjusted models, based on only one criterion, regardless of the nature of the data.

The selection method using the MCS is based on a δ_M equivalence test and an e_M elimination rule. The equivalence test is applied to the set, $M = M^0$. If δ_M is rejected then there is evidence that the models do not have the same predictive power and e_M is used to eliminate the low performance forecast models. This procedure is repeated until δ_M is accepted and \widehat{M}^* is defined for a set of the best models. Using a descriptive level α in all the tests, the method ensures that $\lim_{n \rightarrow \infty} (M^* \subset \widehat{M}^*_{1-\alpha}) \geq 1 - \alpha$. When \widehat{M}^* contains just one model, there is a strong case, in which $\lim_{n \rightarrow \infty} (M^* = \widehat{M}^*_{1-\alpha}) = 1$.

The MCS also generates descriptive levels for each compared model. For a given model $i \in M^0$, descriptive level \widehat{p}_i is the threshold for which $i \in \widehat{M}^*_{1-\alpha}$, if, and only if, $\widehat{p}_i > \alpha$. Therefore, it is improbable that a model with a low descriptive level will be found among the best in terms of forecast accuracy.

The algorithm of the MCS is based on the following steps: (i) initially, make $M = M^0$; (ii) test the hypothesis $H_{0,M}$ using δ_M at descriptive level α .; (iii) if $H_{0,M}$ is accepted, then $\widehat{M}^*_{1-\alpha} = M$ is defined. If not, use e_M to eliminate the low performing model and repeat the routine.

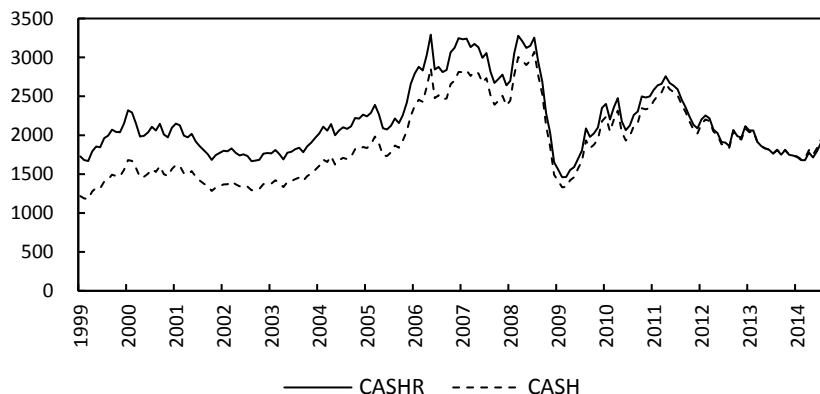


Figure 1 - Aluminum cash price series at nominal values deflated by the CPI (US\$/t)

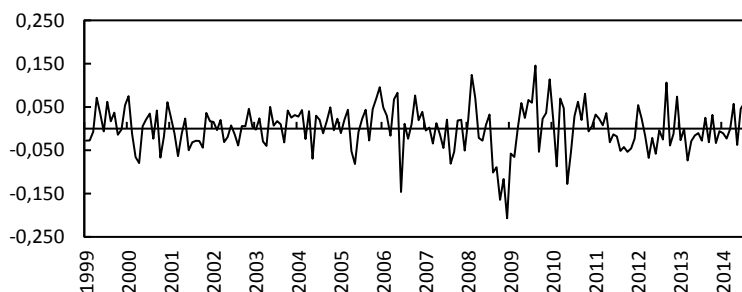


Figure 2 - Transformed Series of the Return on the Price of Aluminum

DATA TREATMENT

ALUMINUM CASH PRICE SERIES

The original data set comprises a monthly aluminum cash price series at nominal values, ranging from January 1999 to September 2014, obtained from the London Metal Exchange (LME). In order to eliminate the inflation effect, the original series (CASH) was deflated using the US Consumer Price Index (CPI), due to the relevance of the North American market within the global context and over the period analyzed, thus obtaining the CASHR series. Figure 1 shows the CASH and CASHR series at constant 2013 values. It was utilized the log return of the original series to make it stationary.

SAMPLES AND EXPLANATORY VARIABLES

The entire sample, from January 1999 to September 2014, was used for selecting the individual models. Each individual model was subsequently adjusted, taking a rolling window of 36 months as the estimation period.

The sample used for adjusting the parameters of each individual model was a 36-month rolling window, considering that the last three years constitute a relevant period for the prospects of the aluminum industry over the subsequent year. Figure 8 illustrates the selection of samples for adjusting the individual models. The first sample was from January 1999 to December 2001. The second sample was from February 1999 to

January 2002. The final sample (the most recent, which was number 142) was from October 2010 to September 2013. Thus, 142 samples were used for adjusting each individual model.

Finally, 9 combined models were used for the final sample, from October 2010 to September 2013. For each of the individual and combined model adjustments, the 12-month ahead

Table 1 – Potential explanatory variables

Category	Description	Measure	Label
Prices	3-month forward aluminum price, deflated by US Consumer Price Index	\$/t	M3R
	Brent oil price	\$/barrel	BRENT
	US Consumer Price Index	Index	CPI2013
	Non-ferrous metals price	Index	NFINDEX
	Price index for industrial non-ferrous products in the US	Index	USPPINFM
Industrial production	Industrial production index in the US	Index	USIP
	Primary metals production in the US	Index	USPMIP
	Industrial production index in China	Index	CHIP
	Industrial production index in Europe	Index	EURIP
	Industrial production index in Japan	Index	JAPIP
Primary aluminum production	Global primary aluminum capacity	t	CAP
	Global primary aluminum production	t	GLPROD
	Global primary aluminum utilization rate	%	GLCAPUT
Primary aluminum consumption	Global primary aluminum consumption	t	GLCONS
	Global primary aluminum consumption, excluding China	t	EXCONS
Primary aluminum inventories	Global average inventory days	days	GLTOTDOC
	Reported global inventory days, including in-transit metal	days	GLREPDOC
	Global inventory days, excluding China	days	EXREPDO
	London Metal Exchange inventory days	days	EXLMEDOC
	Global primary aluminum inventories	days	GLSTOCK
	Reported global inventory, including in-transit metal	t	GLREPSTO
	Global inventory inventory, excluding China	t	EXTOTSTO
	Reported global inventory inventory including in-transit metal	t	EXREPSTO
	London Metal Exchange inventory	t	LMESTOCK
Exchange rate	Effective US exchange rate based on currency basket	Index	FXUSD
	Effective Euro exchange rate	Index	FXEUREFF
	Japan Yen exchange rate	Index	FXJAP
Interest rates	Three month Libot	% p.a.	LIBOR
Sctock market	Dow Jones index	Index	DJ
	Standard-Poors Index (SP 500)	Index	SP500

Table 2 – Total transformations for explanatory variables

Transformation type >	Level	Difference	Total
Potential explanatory variables			30
Transformation by each variable	13	24	37
Total number of transformations	390	720	1.110
Total analyses for weak exogeneity and causality			1.140

For each of the 142 samples, 9 model combinations were used. For example, 9 combined models were adjusted for the first sample, from January 1999 to December 2001.

aluminum price forecast was taken under a point to point basis. In this way, we obtained 142 out-of-sample aluminum price forecasts.

We initially used 30 explanatory variables for the SARIMAX models (Table 1). Such variables were classified into 8 groups: (1) prices; (2) industrial production; (3) production; (4) demand; (5) inventories; (6) exchange rate; (7) interest rates;

weak stationarity of the dependent and independent variables; (b) the weak exogeneity of the independent variables in relation to the dependent variable, in other words, the conditional expectation of error in relation to the

Table 3 – Explanatory variable used in model estimations

Category	Original variable	Transformed variable	Transformation
Prices	M3R	DLM3R_lag4	4-month difference of log price
	BRENT	BRENT_CHG6	6 month percent change
	CPI2013	DLCP2013_lag3	3-month difference of log price
	NFINDEX	NFINDEX_CHG5	5 month percent change
	USPPINFM	DLUSPMIP_lag5	5-month difference of log price
Industrial production	CHIP	CHIP_CHG1	One month percent change
	JAPIP	JAPIP_CHG5	5 month percent change
Primary aluminum inventories	GLSTOCK	GLSTOCK_CHG6	6 month percent change
	EXTOTSTO	EXTOTSTO_CHG6	6 month percent change
Exchange rate	FXUSD	FXUSD_dif4	6 4month percent change
Stock markets	DJ	DJ_dif1	One month difference
	SP500	SP500_dif1	One month difference

(8) stock prices. It was performed 1,140 variable transformations (Table 2), that were then tested for weak exogeneity and Granger causality to arrive at 13 covariates to be used in model estimations (Table 3).

independent variables is zero; (c) the model error is white noise, in other words, it follows a normal distribution, with a zero mean and constant variance.

RESULTS

INDIVIDUAL MODELS

Five individual models were estimated: (i) AR(1) for benchmarking; (ii) ARMA(1,10); (iii) ARMAX 1; (iv) ARMAX 2; (v) Structural model.

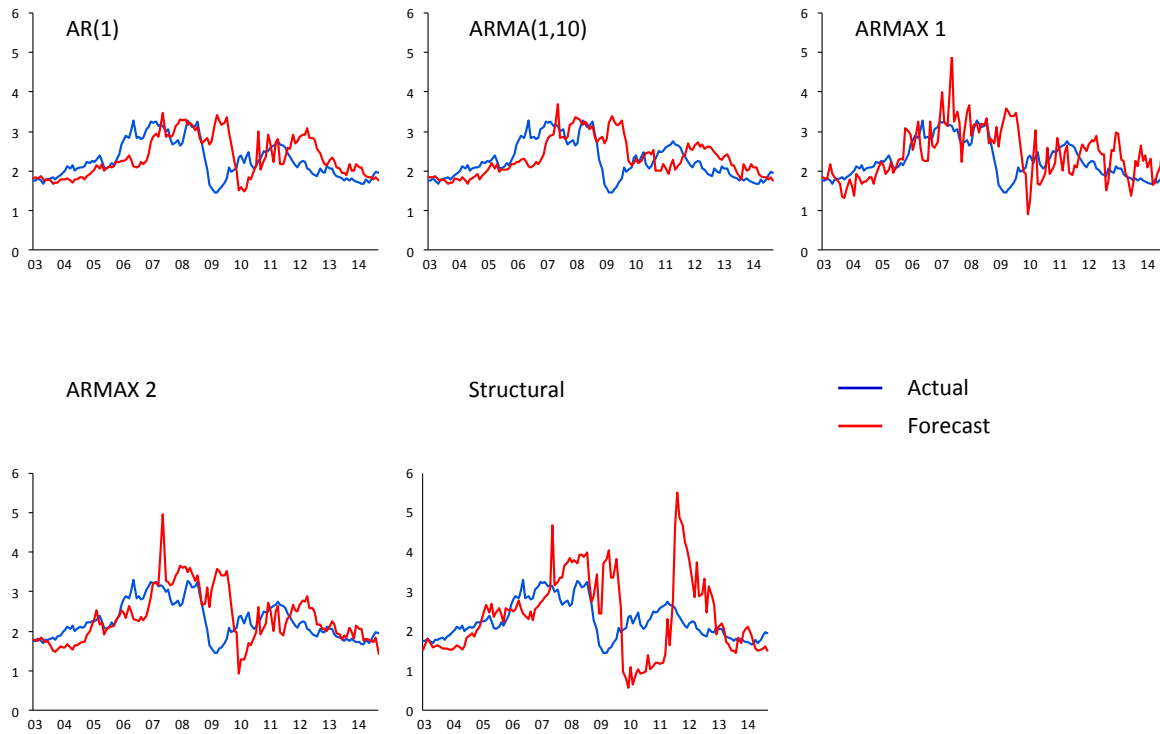
For the two SARIMAX models the following suppositions were tested in their specifications (LUTKEPOHL, 2007; WOOLDRIDGE, 2010): (a) the

For the ARMAX 1 model, two independent variables were used, after performing the tests above: (a) BRENT_dif12: the difference of 12 Brent crude oil price lags; (b) NFINDEX_dif4: the difference of 4 non-ferrous metals index lags. For the ARMAX 2 model, the independent variables used were: (a) DLM3R: the log difference of the 3-month forward price of aluminum; (b) DLGLSTOCK: the log difference of the world stock of aluminum.

Models were estimated using 12-month ahead point to point forecasts, through a 36-month

rolling window. Actual and estimated results are indicated in Figure 3.

Figure 3 – Individual model estimation results



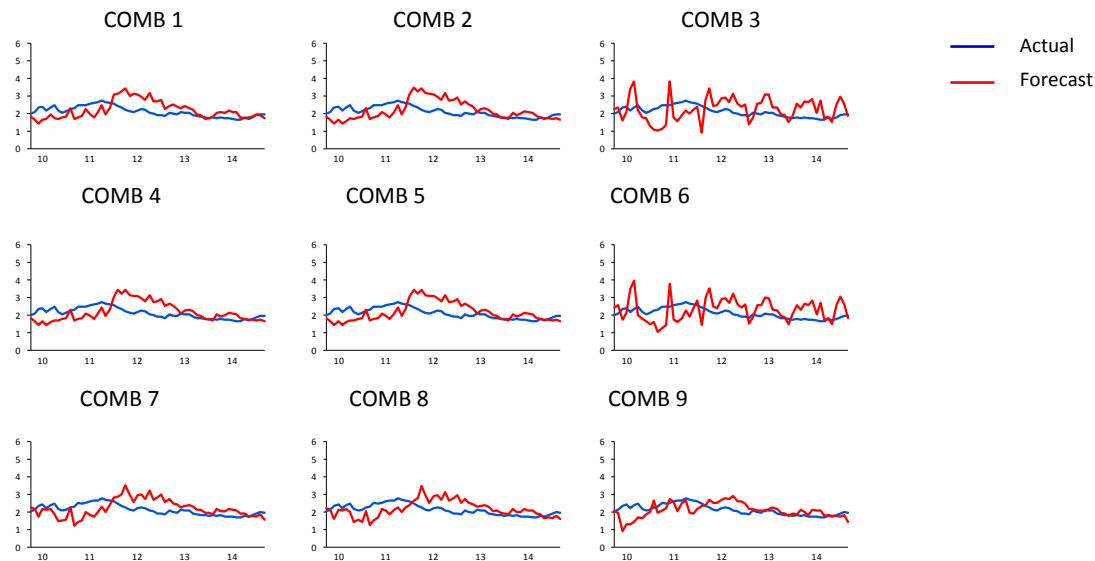
6.2 COMBINED MODELS

For each one of the 60 rolling estimation windows, which resulted in 12-month ahead forecasts, from October 2009 to September 2014, 9 combined models were developed. This represented 540 estimations, with which, added to the 710 estimations of the individual models, a total of 1,250 estimations were produced. The descriptions of the combined models are shown in Table 4. Figure 4 indicates results of actual and forecast values from combined models.

Table 4 – Description of combined models

Model	Method for determining forecast combinations weights
1. COMB_AVG1	Simple arithmetic average of the AR(1), ARIMA, ARMAX1 and Structural models
2. COMB_AVG2	Simple arithmetic average of the AR(1), ARIMA and Structural models
3. COMB_OLS	Ordinary least squares of the AR(1), ARIMA, ARMAX1 and Structural models
4. COMB_RMSE	Weighted mean, based on the inverse of the RMSE of the AR(1), ARIMA and Structural
5. COMB_AICC	Weighted mean based on the AIC of the AR(1), ARIMA, ARMAX1 and Structural models
6. COMB_ERLS	Minimization of the sum of the squares of the residuals, with the restriction that the sum of the weights is 1 for AR(1), ARIMA, ARMAX1 and Structural models
7. COMB_NERLS	Minimization of the sum of the sum of the squares of the residuals, with the restriction that the sum of the weights be 1 and not negative for AR(1), ARIMA and Structural models
8. COMB_NRLS	Minimization of the sum of the sum of the squares of the residuals, with the restriction that the sum of the weights be not negative for AR(1), ARIMA and Structural models
9. COMB_ARMAX	Least ordinary squares of the ARIMA and ARMAX2 models

Figure 4 – Combined model estimation results



COMPARISON OF MODELS

Traditional measures

For the 36 latest months (October/2011 to September/2014), combination model 9 – COMBARMAX - gave the best predictive performance for all traditional measures, followed by ARMAX2 and ARMA. Traditional measures included Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Average Error (MAE), Mean Square Percent Error (MSPE), Root Mean Square Percent Error (RMSPE). Combination 9 is obtained by the OLS of the ARMA and ARMAX2 models. For the 36 out-of-sample latest months (October/2010 to September/2014), the best model was ARMAX2, followed by Combination 9 and ARMA. For the 60 latest months (October/2009 to September/2014), there is an inversion and the ARMA performs best, followed by ARMAX2 and Combination 9. Tables 5, 6 and 7 compare the predictive performance of the models using traditional measures.

On the other hand, the width of the confidence interval of ARMA is very high (around 1,400, 2,000 and 2,600 for 36, 48 and 60 months, respectively) when compared with Combination 9 and ARMAX2. The interval widths of the Combined 9 and ARMAX2 models are similar and the smallest of all the models, at around 46, 76, 88 (US\$/t), respectively for 36, 48 and 60 months.

The two other models include Combination 5 and AR (1). Combination 5 is obtained using the Akaike information criterion method, including

the AR(1), ARMA, ARMAX1 and Structural models. Combination 5 is the fourth best model for the latest 36 and 60 months, followed by AR (1). For 48 months, the opposite happens, with AR(1) being the fourth best model, followed by Combination 5. The AR(1) model, which was accepted as the benchmark, therefore, was the fourth best model in the best case scenario, when performance over the last 48 months is assessed.

It is worth emphasizing that over the last 36 months Combined Model 9 reduced the traditional error measures, including MSE, RMSE, MAE and MAPE, with regard to the values of ARMA and ARMAX2. Based on MSPE, Model 9 and ARMAX2 gave the same performance and, just based on RSPE, ARMAX2 has a negligible advantage. The improvement in the performance generated by Combination 9 is associated with the concept of risk diversification, which was mentioned previously in Section 2. Another important aspect was the introduction of the two exogenous variables in ARMAX2, which were: (i) global aluminum stocks, and (ii) the 3-month forward price of primary aluminum. Days in inventory is a fundamental variable in the global aluminum industry and acts as an indicator of price levels.

Comparison of models using the Model Confidence Set

Table 8 compares the predictive performance of the models by way of the MCS for the last 60, 48 and 36 months of the sample.

Table 5 - Comparison of the Models – 36 latest months

Model	MSE	RMSE	MAE	MAPE	MSPE	RMSPE	CI Amplitude
COMB 9	71,135.54	266.71	163.59	0.0832	0.0180	0.1342	46.98
ARIMAX 2	71,277.47	266.98	163.81	0.0833	0.0180	0.1343	46.41
ARMA	89,531.99	299.22	204.89	0.1055	0.0237	0.1538	1,415.12
COMB 5	124,176.27	352.39	237.20	0.1204	0.0310	0.1760	984.25
AR(1)	134,467.77	366.70	233.77	0.1190	0.0338	0.1839	1,600.15
ARIMAX 1	150,370.41	387.78	263.33	0.1356	0.0398	0.1996	858.03
COMB 8	162,205.55	402.75	258.90	0.1311	0.0400	0.2000	1,370.05
COMB 1	168,288.02	410.23	247.04	0.1237	0.0403	0.2007	969.58
COMB 7	175,182.55	418.55	263.06	0.1329	0.0430	0.2073	1,407.42
COMB 4	183,683.90	428.58	255.25	0.1278	0.0442	0.2102	1,108.55
COMB 2	185,268.43	430.43	255.61	0.1279	0.0445	0.2109	1,107.92
COMB 3	260,321.80	510.22	338.44	0.1751	0.0702	0.2650	2,033.84
COMB 6	263,861.42	513.67	339.07	0.1753	0.0708	0.2661	1,966.39
Structural	487,983.25	698.56	384.84	0.1903	0.1119	0.3345	2,332.31

Table 6 - Comparison of the Models – 48 latest months

Model	MSE	RMSE	MAE	MAPE	MSPE	RMSPE	CI Amplitude
ARIMAX 2	103,882.99	322.31	227.77	0.1079	0.0228	0.1510	66.98
COMB 9	103,909.72	322.35	227.69	0.1079	0.0228	0.1509	68.34
ARMA	133,436.80	365.29	290.72	0.1385	0.0301	0.1735	1,996.53
AR(1)	151,691.27	389.48	283.13	0.1382	0.0364	0.1907	2,441.62
COMB 5	155,459.46	394.28	308.68	0.1482	0.0357	0.1889	1,453.90
ARMAX 1	186,738.60	432.13	334.44	0.1631	0.0452	0.2126	966.57
COMB 8	230,697.47	480.31	365.82	0.1727	0.0504	0.2245	1,899.86
COMB 1	244,254.78	494.22	362.38	0.1689	0.0521	0.2283	1,432.48
COMB 7	267,624.95	517.32	388.50	0.1820	0.0572	0.2391	1,898.55
COMB 4	269,084.75	518.73	378.06	0.1760	0.0576	0.2399	1,590.90
COMB 6	273,061.31	522.55	379.88	0.1767	0.0583	0.2414	1,594.63
COMB 2	413,132.64	642.75	498.68	0.2380	0.0940	0.3066	2,587.16
COMB 3	439,887.15	663.24	509.21	0.2421	0.0983	0.3136	2,802.29
Structural	1,021,803.96	1,010.84	673.29	0.3045	0.1977	0.4446	3,155.47

Table 7 - Comparison of the Models – 60 latest months

Model	MSE	RMSE	MAE	MAPE	MSPE	RMSPE	CI Amplitude
ARMA	143,099.83	378.29	328.17	0.1556	0.0322	0.1793	2,614.87
ARIMAX 2	206,498.50	454.42	342.84	0.1577	0.0416	0.2039	87.60
COMB 9	207,764.76	455.81	343.49	0.1580	0.0418	0.2045	89.22
COMB 5	210,630.33	458.94	392.73	0.1844	0.0454	0.2132	1,836.03
AR(1)	213,865.76	462.46	375.73	0.1784	0.0479	0.2189	3,057.61
ARIMAX 1	272,712.62	522.22	436.45	0.2071	0.0607	0.2463	1,063.25
COMB 8	288,587.58	537.20	454.56	0.2122	0.0618	0.2487	1,330.19
COMB 1	299,584.53	547.34	455.91	0.2100	0.0625	0.2500	1,753.38
COMB 7	323,744.74	568.99	472.77	0.2195	0.0682	0.2611	2,332.36
COMB 4	335,272.77	579.03	480.28	0.2209	0.0699	0.2645	1,918.97
COMB 2	340,748.18	583.74	483.51	0.2222	0.0710	0.2664	1,923.52
COMB 6	545,235.78	738.40	638.69	0.3003	0.1196	0.3459	3,350.75
COMB 3	575,749.72	758.78	648.18	0.3037	0.1246	0.3531	3,570.85
Structural	1,366,861.49	1,169.13	931.70	0.4198	0.2658	0.5155	3,530.25

Table 8 - Comparison of the models using the MCS

60 Months		48 Months		36 Months	
ARMA	1.0000*	ARMAX 2	1.0000*	COMB_ARMAX	1.0000*
ARMAX 2	0.2884*	COMB_ARMAX	0.8345*	ARMAX 2	0.1784**
COMB_ARMAX	0.2873*			ARMA	0.1784**
COMB_AIC	0.2334*				
AR(1)	0.2168*				
ARMAX 1	0.1063**				
Estrutural	0.1063**				
COMB_AVG1	0.1063**				
COMB_AVG2	0.1063**				
COMB_OLS	0.1063**				
COMB_RMSE	0.1063**				
COMB_NERLS	0.1063**				
COMB_ERLS	0.1063**				
COMB_NRLS	0.1063**				

For the latest 36 months (October/2011 to September/2014), Combination 9 (COMB_ARMAX, which includes ARMA and ARMAX2 by the OLS method) gave the biggest MCS p-value (1000), and was by far and away the best model at a significance level of 0.10. The two other comparable models, below Combination 9, were ARMAX2 and ARMA with an MCS p-value of 0.1784.

For the latest 48 months (October/2010 to September/2014), the set of models comparable to the MCS descriptive level of 0.10 included ARMAX2 and COMB_ARMAX. ARMAX gave the best descriptive MCS level, followed by the combination, with a descriptive MCS level of 0.8345.

For the latest 60 months (October/2009 to September/2014), the MCS included 5 models with a forecast performance equivalent to a confidence level of 0.10 of the MCS, as indicated below. All the other models had a weaker performance, with a descriptive MCS level of 0.1063. The models are: ARMA, ARMAX 2, COMB_ARMAX 2, COMB_AIC, AR(1)

The ARMA/ARMAX (COMB_ARMAX) combination was the only model that was among the best sets evaluated by the MCS for the three periods

analyzed. For the three latest years, this combination was the model with the biggest predictive performance; no other performed comparably at a descriptive MCS level of 0.10.

This fact indicates that the introduction of the exogenous variables of the 3-month forward price of aluminum and global stocks levels of primary aluminum was a differential for improving accuracy.

CONCLUSION

The contributions of this work include three main aspects: first, the use of combinations of forecasts and model selection using the Model Confidence Set (MCS), which was not found in the literature on aluminum price forecasts. This approach represents an innovative study in the global aluminum industry, leading to improvements in accuracy and model selection, when compared with traditional methods.

Second, two ARMAX models were proposed, whose approach was also not found in the literature for forecasting aluminum prices. The study showed the relevance of considering the 3-month forward price and world stocks for increasing forecast accuracy.

Third, the work showed that the optimal combination for the 12-month ahead forecast for aluminum prices was the model that included ARMA and ARMAX, adjusted by the least ordinary squares. This combination gave a small confidence interval of below US\$ 100/t, equivalent to 5% of the global average of the entire sample analyzed. This conclusion is relevant when it comes to decision-making by producers (investment decisions), consumers (raw materials supply strategies), participants in the finance market (financing aluminum stocks) and governments (industrial policy).

Suggestions for future studies and for extending this work include:

Combining models that focus on the probability densities of aluminum prices in order to estimate the probability that prices will remain above or below a certain level, or within a band, which is relevant when it comes to the industry taking strategic and tactical decisions

Combination models for high frequency data capable of increasing the accuracy of price forecasts in relation to conventional volatility methods.

Determining prior indicators, based on combinations of independent variables for increasing the predictability of structural shocks, the formulation of scenarios and for supporting decision-making by governments and sectors of the aluminum chain.

REFERÊNCIAS

Bates JM, Granger CWJ. The combination of forecasts. *Operations Research Quarterly*, (20): 451–68, 1969.

Box G, Jenkins G, Reinsel G. *Time Series Analysis*. Wiley, 2008.

Bruggemann R, Lutkepohl H. Forecasting contemporaneous aggregates with stochastic aggregation weights. *International Journal of Forecasting*, Elsevier, 29 (1): 60-8, 2013.

Clark T, Mccracken. Averaging Forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*, 6 (12), 2009.

Cox DR. Statistical Significance Tests. *British Journal of Clinical Pharmacology*, 14: 325-31, 1982.

Christoffersen P, Diebold, F. Optimal prediction under asymmetric Loss. *Economic Theory*, 13: 808-17, 1997.

Diebold F, Lopez J. Forecast combining with neural networks. *Journal of Forecasting*, 15: 49-61, 1996.

Diebold FX, Mariano RS. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13: 253–63, 1995.

Driehuis W. *Primary commodity prices: analysis and forecasting*. Rotterdam: Rotterdam University Press, 1976.

Duncan RC. *The outlook of primary commodities*. Staff Working Paper n. 11. Washington: World Bank, 1984.

Dooley G, Lenihan H. An assessment of time series methods in metal price forecasting. *Resources Policy*, (30): 208–17, 2005.

Elliott G, Timmermann A. Optimal Forecast Combinations Weights under Regime Switching. *International Economic Review*, (46): 1081-1102, 2005.

Engle R, Granger C, Kraft D. Combining competing forecasts of inflation using a bivariate arch model. *Journal of Economic Dynamics and Control*, (8): 151-65, 1984.

Fan WG, Deng F. Variable weight combining forecasts based on forecasting error. *International Conference on Control and Automation*, 5: 1610-3, 2007.

- Figlewski S. Optimal aggregation of money supply forecasts: accuracy, profitability and market efficiency. *Journal of Finance*, (28): 695-710, 1983.
- Granger CW, Newbold P. *Forecasting Economic Time Series*. Academic Press: San Diego, 1986.
- Granger CW, Newbold P, Ramanathan R. Improved methods of combining forecasts. *Journal of Forecasting*, (19): 537-60, 1984.
- Hansen P, Lunde A, Nason J. The Model Confidence Set. *Econometrica*, 79 (2): 453-97, mar. 2011.
- Harvey A. *Forecasting: structural time series models and the kalman filter*. Cambridge: University Press, 1989.
- _____, Jaeger A. The effects of seat belt legislation on british road casualties: a case study in structural time series modeling. *Journal of the Royal Statistical Society*, (149): 187-227, 1993.
- Hendry D, Clements M. Pooling of forecasts. *Economics Journal*, Elsevier, (5): 1-26, 2002.
- Hsiao C, Wan SK. Is there an optimal forecast combination? *Journal of Econometrics*, (178): 294-309, 2014.
- Issler J, Lima R. A panel data approach to economic forecasting: the bias-corrected average forecast. *Journal of Econometrics*, 152 (2): 153-64, 2009.
- Kim JRK. Combining forecasts using optimal combination weight and generalized autoregression. *Journal of Forecasting*, 27: 419-32, 2008.
- Kolassa S. Combining exponential smoothing forecasts using Akaike weights. *International Journal of Forecasting*, 27 (2): 238-51, 2011.
- Kouassi E, Labys W, Colyer D. *The structural time series approach to forecasting commodity prices*. Morgantown: West Virginia University, 1998.
- Labys W, Kouassi E. Cyclicity, stationarity and the structural forecasting of commodity prices, working paper. *Natural Resource Economics Program*, West Virginia University, 2004.
- _____, Pollack PK. *Commodity models for forecasting and policy analysis*. New York: Nichols Publishing, 1984.
- Lam K, Mui H, Yuen M. Note on minimizing absolute percentage error in combined forecasts. *Computers & Operations Research*, 8 (11): 1141-7, 2001.
- Leamer E. Let's take the con out of econometrics. *American Economic Review*, (73): 31-43, 1983.
- _____. *Specification searches: ad hoc inference with nonexperimental data*. New York: Wiley, 1978.
- Leuthold RM, McCormick A, Schmitz A, Watts DG. Forecasting daily hog prices and quantities: a study of alternative forecasting techniques. *Journal of the American Statistical Association*, (65): 65-90, 1970.
- Liang KY, Lee JC, Shao KSH. On the distribution of the inverted linear compound of dependent f-variates and its application to the combination of forecasts. *Journal of Applied Statistics*, (33): 961-73, 2006.
- Lo A, Mackinley C. Data snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3: 431-68, 1990.
- Lovell MC. Data Mining. *Review of Economics and Statistics*, 45: 1-12, 1983.
- Makridakis S, Winkler R. Averages of forecasts: some empirical results. *Management Science*, (29): 987-96, 1983.
- _____, _____. The accuracy of extrapolation methods: results of a forecasting competition. *J. Forecasting*, 1 (11): 1-153, 1982.
- Mayer T. Economics as a hard science: realistic goal or wishful thinking? *Economic Inquiry*, 18: 165-78, 1980.
- Miller Júnior RG. *Simultaneous statistical inference*. New York: Springer-Verlag, 1981.
- Mills T. *Modeling trends and cycles in economic time series*. New York: Palgrave, 2003.

Nelson CR. The prediction performance of the FRB-MIT-PENN model of the us economy. *American Economy Review*, (62): 902-17, 1972.

Newbold P, Granger CWJ. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, Series A*, (137): 131-46, 1974.

Pesaran MH, Timmermann A. *Selection of estimation window in the presence of breaks*. San Diego: Department of Economics, University of California, 2005.

Ren Y, Ru Y. Application of ARMA model in forecasting aluminum price. *Applied Mechanics and Materials*, (155-156): 66-71, 2012.

Slutsky E. The summation of random causes as a source of cyclic processes. *Econometrica*, (5): 105-46, 1927.

Smith J, Wallis K. Simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71 (3): 331-55, 2009.

Stoffer D, Schumway R. *Time Series Analysis and Its Applications*. [s.l]: Springer, 2006.

Stock JH, Watson M. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle RF, White H. (eds.). *Cointegration, Causality, and Forecasting A Festschrift in Honour of Clive W. J. Granger*. Cambridge University Press, 2001.

_____, _____. Combination of forecasts of output growth in a seven-country dataset. *Journal of Forecasting*, (23): 405-30, 2004.

White H. Reality check for data snooping. *Econometrica*, 68 (5): 1097-126, set. 2000.

Yule GU. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions*, (22-A), 1927.