DOI: https://doi.org/10.14393/REE-v39n1a2024-66859
Instituto de Economia e Relações Internacionais – Universidade Federal de Uberlândia

economia
ensaios

Modeling and Forecasting of Industrial Production in Brazil

Modelagem e Previsão da Produção Industrial no Brasil

Lucélia Pontes e Pontes[1]

Carlos Enrique Carrasco-Gutierrez[2]

Abstract: Industrial Production is considered a relevant measure for analyzing the economic situation and making decisions within a country. In this study, we propose several alternative short-term forecasting models for the Industrial Production series in Brazil. In our analysis, we consider the univariate ARIMA model, the dynamic distributed lag ARDL model, and the multivariate VAR model. Additionally, we incorporate various methods of forecast combination in the final selection to enhance results. Within the study, we integrate aggregate variables such as the Selic interest rate, Bovespa index, energy consumption, revenue, working hours, imports of machinery and equipment, employment, and inflation rate. The findings indicate that the ARDL model exhibited the best forecasting performance for horizons of 1, 3, 6, and 12 steps ahead. However, in comparison to forecast combination methods, the OLS-AVG model demonstrated superior outcomes, underscoring that diversifying forecasts leads to a reduction in diversifiable error.

Keywords: ARDL; Forecasting; Industrial Production; Model Combinations; VAR.
JEL Classification: C51; C52; C53; E23

Resumo: A Produção Industrial é considerada uma medida relevante para analisar a conjuntura econômica e na tomada de decisões em um país. Neste trabalho, propomos várias alternativas de modelos de previsão de curto prazo para a série da Produção Industrial no Brasil. Na análise, consideramos o modelo univariado ARIMA, o modelo dinâmico de defasagens distribuídas ARDL e o modelo multivariado VAR. Além disso, incluímos diferentes métodos de combinação de previsões na escolha final, visando aprimorar os resultados. No estudo, incorporamos as variáveis agregadas, como taxa de juros Selic, índice da Bovespa, consumo de energia, faturamento, horas trabalhadas, importações de máquinas e equipamentos, emprego e taxa de inflação. Os resultados indicam que o modelo ARDL apresentou o melhor desempenho de previsão para horizontes de 1,3, 6 e 12 passos à frente. No entanto, em comparação com os métodos de combinação de previsões, o modelo OLS-AVG demonstrou melhores resultados, mostrando que a diversificação das previsões leva à redução do erro diversificável.

Palavras-chave: ARDL; Previsão; Produção Industrial; Combinação de Modelos; VAR.
Classificação JEL: C51; C52; C53; E23

[1] PhD in Economics from the Catholic University of Brasília (UCB) and Professor at Faculdade La Salle Manaus. E-mail: lucelia.pontes@lasalle.org.br. ORCID: https://orcid.org/0000-0002-9272-7340.
[2] Professor and researcher of the Postgraduate Programs in Economics and Public Policies at the Catholic University of Brasília (UCB).E-mail: carlosenrique@p.ucb.br. ORCID: http://orcid.org/0000-0003-2215-7730.

1. Introduction

The importance of industrial activity in the economy is due to the fact that industrial development represents one of the main links between job creation and income, given that production is seen as the foundation of a country's entire economic system (LAPO, 2014; HERRIGEL, 2015). Industries are considered the strategic focus of those countries that wish to overcome their situation of economic dependence.

Industrial production (IP) is recognized as one of the most important measures of the level of activity and economic process in Brazil. Despite the existence of the Gross Domestic Product (GDP) and the fact that it is considered the main measure of this level, industrial production has a fundamental difference to GDP, as it represents the transformation of raw materials into marketable products. Its continuous monitoring makes it possible to evaluate and compare the performance of the industry sector and, consequently, of the country's economy.

Industrial production forecasting originated in an attempt to solve the need for managers to outline the future of their activities with assertiveness, since the information generated through econometric models is of paramount importance from a governmental point of view, given that predicting the industrial production index is an important issue in short-term economic analysis.

The aim of this paper is to find the short-term forecasts for Industrial Production in Brazil. We considered various forecasting models in the analysis: univariate models, causal models, multivariate models, as well as forecast combination models.

Several authors have studied the forecasting of industrial production. For example, Thury and Witt (1998) studied the industrial production series of the economies of Austria and Germany; Birchenhall et al. (1999) in Germany, France and the UK; Bodo et al. (2000) in the Euro area; Simpson et al. (2001) in the UK; Lupi and Bruno (2003) in Italy; Hassani et al. (2009) for important sectors of the German, French and British economies and; Chiu et al. (2017) for the USA. In Brazil we have the work of Markwald et al. (1989), Hollauer et al. (2008), Feitosa and Tabak (2010) and Zuanazzi and Ziegelmann (2014).

Thury and Witt (1998) presented industrial forecasting models for the countries of Austria and Germany using structural time series models. One of the models applied was the ARIMA and was used as a basis for comparison with basic structural forecasting models. Bodo et al. (2000) proposed a series of alternative short-term forecasting models, ranging from simple ARIMA models to more complex cointegrated VAR and conditional models, to forecast the industrial production index in the Euro area. The results show that the conditional error correction model achieves the best score in terms of predictive ability.

Simpson et al. (2001) defined information through leading indicators for modeling and forecasting the UK quarterly production index using both linear and non-linear models. The main variables include a short-term interest rate, the dividend yield on the stock market and the optimism balance of the CBI's quarterly survey.

Lupi and Bruno (2003) suggested a simple model for forecasting production in Italy up to 6 months ahead, and showed that the forecasts produced using the model outperform the VAR and some popular forecasts, as well as those resulting from an ARIMA model used as a reference and some alternative single equation models.

Hassani et al. (2009) compared the singular spectrum analysis methods, ARIMA and Holt-Winters, and demonstrated that SSA is a very powerful tool for analyzing and forecasting economic data. The SSA outperformed the ARIMA and Holt-Winters methods in predicting the values of the production series according to the RMSE criterion, particularly at horizons of h = 3, 6 and 12 months.

Chiu et al. (2017) estimated a model using US data on industrial production growth, inflation, interest rates and stock returns, the VAR model with t-disturbances results in higher density forecasts for industrial production and stock returns than alternatives that assume Gaussianity.

Few studies have discussed modeling and forecasting industrial production in Brazil. Markwald et al. (1989) discuss industrial production forecasting by working with leading indicators and time series models, with the aim of comparing the leading indicator methodology with structural time series models, in order to forecast cyclical reversals in a target series. The results show that the structural models performed better than the leading indicator models.

Hollauer et al. (2008) used ARMA and VAR models for quarterly frequency data, using the Brazilian industrial GDP as an indicator and the macroeconomic variables; long and short interest rates, inflation and the forecast of industry balances. The authors conclude that the best forecasting model is the VAR model with cointegration. In addition, the predictive combination models, in most cases, outperformed the other models, which already had good predictive capacity.

For Feitosa and Tabak (2010), the model applied was the so-called Wavelet, which worked on the decomposition of the multiresolution to the spread and industrial production, and investigated whether the term spread presents information that can help predict the path followed by industrial production, they also investigated whether all the predictive capacity of the spread is due to monetary policy, and the results obtained suggest that the spread of income contains relevant information and this predictive power varies in time patterns. The results indicate that the predictive power of the spread is not fully explained by monetary policy.

Zuanazzi and Ziegelmann (2014) looked at the MIDAS and UMIDAS models, comparing their out-of-sample forecasting results with the ARMA benchmark for Brazilian quarterly GDP growth, using 16 monthly financial and economic series as potential predictors, covering the second quarter of 1996 to the fourth quarter of 2012, and the results showed good out-of-sample forecasting performance for several individual regressors.

In this work we used time series with a monthly frequency covering the period from January 2002 to January 2017. In addition to industrial production, we incorporated macroeconomic and financial variables that are fundamental to activity in Brazil: the Selic interest rate, inflation and the Bovespa index. The choice of these variables is relatively standard and follows the literature (BANBURA et al., 2010; D'AGOSTINO, GAMBETTI and GIANNONE, 2013; PRIMICERI, 2005) which used equivalent variables when determining small-scale forecasting models. In addition to the variables mentioned above, we added variables not considered in most studies, such as: hours worked, employment, electricity consumption and imports of machinery and equipment, with the aim of capturing some macroeconomic effects that help predict industrial production. The results of this work indicate that the method of combining different forecasting models generated better forecasting results for all the horizons analyzed.

In addition to this introduction, the structure of the article is organized as follows: in chapter 2 we present the econometric methodology; in chapter 3 the databases and the results found in the empirical modeling and; finally, in chapter 4 the conclusions.

2. Methodology

In this study we used various econometric models to forecast the Industrial Production series. We considered ARIMA, ARDL and VAR models, as well as forecast combination methods such as simple averaging (S-AVG), variance-based weighted averaging (VB-AVG) and the ordinary least squares performance model (OLS-AVG). In order to determine the best forecasting model, we followed three steps: identification and estimation; diagnosis of the models and; evaluation of out-of-sample performance. The specifications of each estimated model are defined below.

2.1. ARIMA model (p,d,q)

The peculiarity of univariate models is that they capture the stylized facts of time series that are explained solely by the behavior of the series itself. The ARIMA (p,d,q) model is defined for the time series y_t in the following way:

$$\phi(L)\Delta^d y_t = c + \theta(L)\varepsilon_t \tag{1}$$

where,

$$\Delta^d y_t \sim I(0)$$

and,

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q,$$

$$\Delta = 1 - L$$

$\phi(L)$ and $\theta(L)$ are the polynomials associated with the AR and MA parts, respectively. All the roots of the autoregressive polynomial $\phi(L)$ are outside the unit circle. The $\phi's$ and $\theta's$ are the autoregressive and moving average parameters, respectively. The error component, ε_t, is a Gaussian white noise process, $\varepsilon_t \sim \mathrm{RBN}(0, \sigma^2)$.

2.2. Autoregressive Distributed Lag Model (ARDL)

The dynamic distributed lag model can be used to satisfactorily describe the evolution of the economy and its reactions over time. This approach involves creating a general model that includes a number of independent variables and their lags, as well as considering the lags of the dependent variable. This situation allows us to create ARDL models, which in their generic form can be described as:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \gamma_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q_j} \beta_{ji} \, x_{j,t-i} + \varepsilon_t \qquad (2)$$

We use the standard ARDL specification $(p, q_1, q_2, \dots, q_k)$, where p is the number of lags of the dependent variable, q_1 the number of lags of the first independent variable, and q_k the lag number of the kth independent variable.

2.3. Multivariate Model (VAR)

The vector autoregressive model of order p - VAR(p) - is composed of a vector of *n* endogenous variables $Y_t \in \mathbb{R}^n$ that make up *n* equations structured in:

$$Y_t = c + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \varepsilon_t \qquad (3)$$

where $c \in \mathbb{R}^n$ is the intercept vector, $A_i \epsilon \mathbb{R}^{n \times n}$, $i = 1,2, \dots, p$, are the model coefficient matrices and ε_t is the vector of errors satisfying the following properties: $E[\varepsilon_t] = 0 \in \mathbb{R}^n, E[\varepsilon_t \varepsilon_t^T] = \Sigma$, where Σ is the variance-covariance matrix of the errors. It is possible to incorporate independent variables into model (3) as shown in (4)

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + B' X_t + \varepsilon_t \qquad (4)$$

Where X_t is a vector of exogenous variables and B is a vector of parameters.

2.4. Combining forecasting models

The purpose of combining forecasting models is to build superior performance forecasts. The approach is defined as:

$$Y_t = w_1 \hat{Y}_t^1 + w_2 \hat{Y}_t^2 + \cdots + w_M \hat{Y}_t^M \qquad (5)$$

Where $\hat{Y}_t^j$ corresponds to the forecast obtained by the model j, for j =1,2,..*M*. The parameter w_j, for j =1,2,..*M*, is the weight assigned to the prediction result obtained by the model j. The way the parameter is defined w_j, defines the type of forecast combination model. We'll look at three types of model, S-AVG, VG-AVG and OLS-AVG.

2.4.1 Model S-AVG

The S-AVG model is defined by the simple arithmetic mean which assigns an equal weight to all the models, making the method free of errors in determining the weights.

$$w_j = \frac{1}{M}, \forall\, j \tag{6}$$

where M is the number of prediction models, $w_j \epsilon \mathbb{R}$ is the weight of the jth individual algorithm.

2.4.2 Model VG-AVG

This variance-based weighting model (VB-AVG) assigns weights to the forecast models using the performance of past forecasts. Where $e_t^k = (Y_t - \hat{Y}_t^k)$ the prediction error of the kth model in t, k = 1,...,M. The VB-AVG weights are found using the normalized squared prediction errors:

$$w_k = \left(\frac{\sum_{t=1}^{N} \left(e_t^k\right)^2}{\sum_{t=1}^{N} ((e_t^1)^2 + (e_t^2)^2 + \cdots + (e_t^M)^2)} \right)^1 \tag{7}$$

The VB-AVG prediction errors are calculated using pairs of training and validation data sets included in the sample. This scheme is based on the fact that predictions from models with large sample errors are given less weight and vice versa.

2.4.3 Model OLS-AVG

This model consists of finding the weights using the Ordinary Least Squares (OLS) method. The estimates of the weights are found as follows: the model (8) is estimated using the OLS method and the estimated parameters correspond to the weights of the model $\hat{w}_k, k = 1, \ldots, M$:

$$Y_t = w_1 \hat{Y}_t^1 + w_2 \hat{Y}_t^2 + \cdots + w_M \hat{Y}_t^M + \epsilon_t \tag{8}$$

where ϵ_t is the model error. The estimated equation (11) is used to find the prediction combination $\hat{y}_t$ in *t* using this method.

$$\hat{y}_t = \hat{w}_1 \hat{y}_t^1 + \hat{w}_2 \hat{y}_t^2 + \cdots + \hat{w}_M \hat{y}_t^M \tag{9}$$

3. Results

3.1. Database

The period of analysis runs from January 2002 to December 2017, on a monthly basis. Industrial production data is obtained from the Brazilian Institute of Geography and Statistics (IBGE). Hours worked, staff employed, imports of machinery and equipment, inflation and the Selic interest rate were obtained from IPEA. Electricity consumption is published by Eletrobrás and made available by the Energy Research Company (EPE). The BOVESPA financial market index was obtained by Economatica.

Hours worked (Ht) measures the evolution of production in industries by workers and represents the total hours worked by the personnel employed in production in the local unit surveyed. Personnel employed (Pe) symbolizes the efficiency of people in product manufacturing processes and represents a proxy for the expectations of entrepreneurs. Real turnover (Fr) shows the sum of the gains made by industries from the sale of their products, discounting inflation. Imports of machinery and equipment (Impor) stimulate industrial activity and help increase productivity. Electricity consumption (Cons) describes the amount of energy in units (Gwh) used by industries. It is characterized as a fundamental input and represents the driving force for powering machinery. The inflation rate is obtained from the national consumer price index (IPCA).

In all the models, we considered a time dummy variable in order to capture the effect of the subprime crisis on the economy. The dummy assumes a value of one from December 2008 to December 2009 and zero otherwise.

3.2. Descriptive Statistics

Table 1 shows the main descriptive statistics obtained from the four series selected for the research, providing an overview of the behavior of the average monthly indices over the period studied.

Table 1 – Descriptive statistics of the series used, 2002 - 2017

	Industrial Production	Inflation IPCA	SELIC interest rate	BOVESPA Index	Hours worked
Average	92.66	0.0052	0.010	4.6456.22	92.89
Median	92.45	0.0047	0.0097	51.605.00	94.02
Maximum	112.60	0.302	0.020	76.402.00	109.82
Minimum	69.70	-0.0023	0.0049	8.623.00	73.12
Standard Deviation	9.99	0.0038	0.0032	18.326.50	8.68
Kurtosis	2.22	14.52	3.32	2.19	2.21
Coefficient of Variation	0.10	0.73	0.31	0.394489	0.093
Observations	192	192	192	192	192

	Staff employed	Actual Turnover	Imports of machinery and equipment	Electricity consumption (GWh)
Average	90.83	85.33	91.43	13889.81
Median	91.51	83.90	92.28	14075.29
Maximum	101.31	109.30	106.58	15886.31
Minimum	78.47	64.27	69.26	9705.94
Standard Deviation	7.20	10.49	7.95	1370.96
Kurtosis	1.62	2.26	2.53	3.02
Coefficient of Variation	0.079	0.12	0.086	0.98
Jarque-Bera	15.49	5.25	10.27	22.27
Observations	192	192	192	192

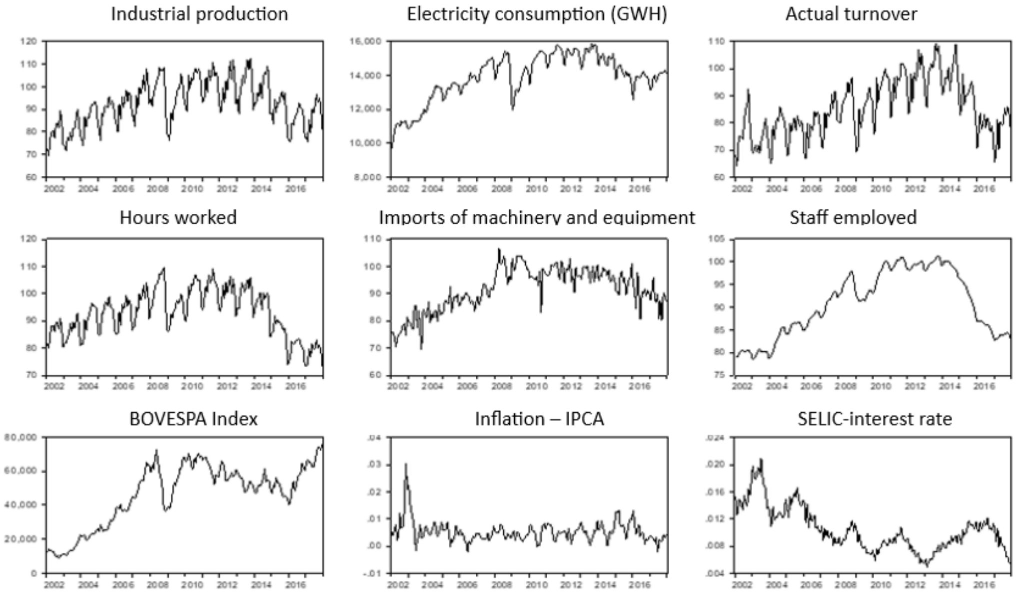
Source: Elaborated by the author.

3.3. Industrial production forecast series

Figure 1 shows the series for the period from January 2002 to December 2017. We can see that there is evidence of seasonality in most of the series presented. To avoid the complexity of the analysis by incorporating seasonality into the models, all the series were deseasonalized using the *Census* ARIMA X-12 procedure.

We can see graphically that the series of electricity consumption, real turnover, hours worked, imports of machinery and equipment, and personnel employed seem to follow the behavior of industrial production. The effects of the subprime crisis can also be seen in the time series.

Figure 1 – Time series used for Brazilian industrial production



Note: Industrial production series for general industry, inflation rate measured by the National Consumer Price Index (IPCA), BOVESPA Index, Selic interest rate.

3.4 Unit root test

In principle, before estimating econometric models with temporal data, it is necessary to check the stationarity of the stochastic processes, which aims to verify the presence of a unit root in the series being worked on. The Augmented Dickey-Fuller (GLS) and Phillips-Perron tests, which are constantly applied in the literature, were used. Table 2 shows the results of the tests for the presence of a unit root for all the series involved.

As you can see, the statistics presented show that Industrial Production, the Logarithm of the Selic Interest Rate, Hours Worked, Real Turnover, Personnel Employed, Electricity Consumption and Imports of Machinery and Equipment have a unit root and the Inflation variable and the Ibovespa Return are stationary. The results of the new unit root tests carried out with the series in first difference indicated that the series also became stationary.

Table 2 – Unit root test for level and difference variable

Variables	Level				1st Difference			
	Dickey Fuller – GLS		Phillips – Perron PP		Dickey Fuller – GLS		Phillips - Perron PP	
	μ	μ e δ	μ	μ e δ	μ	μ e δ	μ	μ e δ
Pi	- 1.93	- 1.64	- 2.23	- 1.98	- 4.95**	- 5.08**	- 20.09**	- 21.03**
Ht	- 1.28	- 1.05	- 1.30	- 1.33	- 3.43**	- 4.49**	- 20.39**	- 20.93**
Fr	- 1.45	- 0.92	- 2.26	- 2.23	- 4.40**	- 21.18**	- 22.82**	- 23.22**
Pe	- 1.69	- 0.37	- 1.31	0.06	- 3.30**	- 3.71**	- 6.53**	- 7.21**
Cons	- 2.82	- 1.95	- 2.76	- 2.02	- 6.56**	- 6.70**	- 13.48**	- 13.65**
Impor	- 2.40	- 1.53	- 3.36	- 3.89	- 23.13**	- 23.17**	- 35.34**	- 42.95**
Ibov	- 5.92**	- 5.99**	- 12.27**	- 12.31**	- 22.58**	- 22.53**	- 63.70**	- 63.90**
Inf	-3.89**	- 3.84**	- 5.65**	- 5.80**	- 16.38**	- 16.34**	- 17.45**	- 17.40**
Selic	- 3.17	- 2.56	- 1.51	- 2.35	- 3.64**	- 3.62**	- 18.18**	-18.14 **

Notes: (1) Applied to test equations with intercept. Significance of 5%. (2) Critical values: -2.578636 (1%). -1.942710 (5%) and -1.615460 (10%). Use the *Modified Schawz* method. (2) We used the *Bastlett Kernel* with *Newey-West Bandwidth* method of estimating. Critical values: -3.466176 (1%). -2.877186 (5%) and -2.575189 (10%). Use the *Modified Schawz* method. Critical values: 0.7390 (1%). 0.4630 (5%) and 0.3470 (10%). Use Default (Bartlett kernel) and Bandwidth Newey-West Bandwidth. The series was considered stationary or non-stationary when three of the tests indicated the same position.

Not least, recent studies argue that there is weakness in the results of the ADF and KPSS tests in the presence of potential structural breaks, showing evidence of non-stationarity. Perron (1989) states that, when considering structural breaks in the series, traditional tests are less able to reject the hypothesis of a unit root that is actually false.

3.5. Strategic modeling

In the procedure to find the best forecasting model for industrial production, we separated the database into two periods, the first for estimating and identifying the models and the second for comparing forecasting performance. The first set of data corresponds to the period from January 2002 to December 2013, leaving the last three years, January 2014 to January 2017, for comparison of the models' forecasting performance.

3.5.1 Univariate model

The Industrial Production series shows a clear trend, and this sign of non-stationarity was confirmed with the unit root tests (integration of order 1, I(1)). Once the variable was stationary, the model was chosen using a set of specifications defined by the parameters p and q. The number of lags of the industrial production variable was estimated by minimizing the Akaike (AIC), Schwarz (SC) and Hannan Quinn (HQ) information criteria using a maximum of 12 lags. The best models chosen according to the information criteria are the ARIMA (0,1,1) and the ARIMA (5,1,5) models. The diagnostic tests show that the ARIMA (5,1,5) model satisfies all the assumptions of the correctly specified model. The results of the diagnostic tests for this model are shown in Table 3.

3.5.2 Model ARDL

The ARDL models were identified using the Akaike (AIC) and Schwarzsc (SC) information criteria, considering a maximum of three lags for the dependent variable (for $p < 3$) and $q_k < 5$ for k=1,2,..,8 which correspond to the other variables.

Three models were selected according to the lowest values of the information criteria, of which only the ARDL model (1,0,4,3,3,2,2,0,0) passed all the diagnostic tests, as shown in Table 3.

3.5.3 Multivariate Model

The Schwarz (SC), Akaike (AIC) and Hannan Quinn (HQ) information criteria are used to identify the VAR model (p). These statistics make it possible to choose the number of lags, p, which defines the VAR model (p), being the model chosen the one with the lowest value of the information criteria. The results presented in the appendix show that the SC and HQ criteria selected p=1, with the AIC choosing p=8 (see Table 5 in appendix). These two specifications were considered, of which only the VAR model (8) passed all the model specification tests, as shown in Table 3.

Table 3 – Specification test statistics for models 2002M01 – 2013M12

Models	ARCH Hesterocedasticity	LM-test Autocorrelation	Jarque-Bera Normality
ARIMA (5,1,5)	1.4417 (0.2289)	0.5868 (0.6269)	4.4890 (0.1059)
ARDL (1,0,4,3,3,2,2,0,0)	2.4729 (0.1164)	1.1043 (0.2613)	0.0537 (0.9734)
VAR (8)	2877.21 (0.2906)	49.9899 (0.06060)	3.6884 (0.1581)

Note: ARCH test, Least Squares method, with lag=1. (2) Breusch-Godfrey Serial Correlation LM Test, method Least Squares,com lag=2. (3) Jaque-Bare test for normality of residuals. The values in brackets are the P-values of the tests (Prob. Chi-Square) and the first value is that of the F-static.

3.6. Predicition Performance

The selection of the best forecasting model was based on the criterion of out-of-sample forecasting performance. We used the database from January 2014 to January 2017 to assess the forecasting performance of the best models identified in the previous phase. As accuracy criteria we used the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) statistics.

Table 4 shows the out-of-sample forecasting performance statistics of the five best models identified for the horizons of one, three, six and twelve steps ahead. The forecasting results of the VB-AVG and OLS-AVG models were estimated using combinations of forecasts from the best ARIMA, ARDL and VAR models. The S-AVG model had the worst performance among the forecast combination models and the results have not been reported in this table.

According to the results obtained through the RMSE, MAE and MAPE statistics, the model with the best forecasting performance, for a horizon of 1, 3, 6 and 12 steps ahead, is the OLS-AVG, the second best model is the dynamic ARDL distributed lags model, and the third model is the multivariate VAR model.

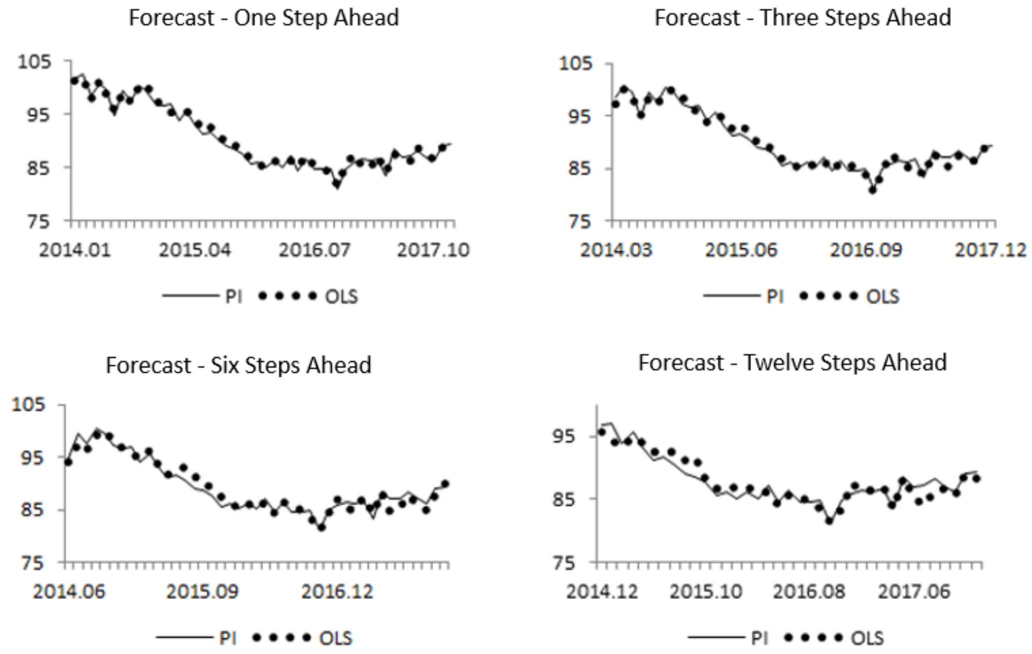
Table 4 – Predictive performance of the estimated models

Models	Horizon forecasts (*h*)			
	1	3	6	12
	RMSE			
ARIMA	2.74	5.49	10.23	15.85
ARDL	0.93**	1.10**	1.28**	1.61**
VAR	0.98***	1.32***	2.12***	4.19***
VB-AVG	2.23	2.32	3.84	7.32
OLS-AVG	0.78*	0.94*	1.15*	1.30*
	MAE			
ARIMA	2.30	5.08	9.81	15.52
ARDL	0.71**	0.87**	1.07**	1.30**
VAR	0.79***	1.08***	1.62***	3.39***
VB-AVG	1.87	1.91	2.97	6.12
OLS-AVG	0.65*	0.79*	0.93*	1.02*
	MAPE			
ARIMA	2.59	5.74	11.15	17.80
ARDL	0.79**	0.96**	1.20**	1.47**
VAR	0.88***	1.21***	1.83***	3.88***
VB-AVG	2.10	2.11	3.30	6.95
OLS-AVG	0.72*	0.86*	1.04*	1.29*

Notes: *Best model, ** second best model, and, *** third best forecasting model for horizons of 1, 3, 6 and 12 steps ahead. The ARIMA model corresponds to ARIMA (5,1,5), the ARDL model to ARDL (1,0,4,3,3,2,2,0,0), and the VAR model is VAR (8).

Figure 2 shows the Industrial Production Forecast for the period from January 2014 to December 2017 for the horizons of 1, 3, 6 and 12 steps ahead, for the OLS-AVG model.

Figure 2 – Observed values of industrial production and values predicted by the model



4. Conclusions

This paper aims to find a short-term model of industrial production in Brazil. Various econometric models for time series data were considered in this study, ranging from univariate ARIMA models to multivariate VAR models, as well as the dynamic ARDL distributed lag model and forecast combination models.

The empirical strategy separated the database into two sets, training and test. The training set was used to identify, estimate and validate the models; and the test set was used to compare the out-of-sample forecasting performance, based on the accuracy of the RMSE, MAE and MAPE forecasting statistics. Among the best models identified in the first phase, the forecasting results showed that the OLS-AVG forecast combination model generated better results for horizons of 1,3, 6 and 12 steps ahead, showing that diversifying forecasts leads to a reduction in the diversifiable error. The second model with the best forecasting performance was ARDL and the third best model was VAR. The ARDL and VAR models perform better than the classic univariate ARIMA model, which is commonly used in the literature as the benchmark model. It is clear that the inclusion of more variables and specification information in the models and their lags provide important information on predictive capacity.

Acknowledgments

Lucélia acknowledges the support and contributions of the professors from UCB and Faculdade La Salle Manaus, with special thanks to Professor and Academic Director Fátima Durães. Carlos Enrique thanks CNPq and FAPDF for their support in the development of this research.

References

BIRCHENHALL, C.R., HERAVI, S., & OSBORNA, D.R. Seasonal unit roots and forecasts of two-digit European industrial production. **International Journal of Forecasting,** 1999, Elsevier .

BODO, G.,GOLINELLI, R,. & PARIGI, G. Forecasting industrial production in the Euro área, **Empirical Economics**, Volume 25, Number 4, 2000

BRUNO,G., LUPI, C. Forecasting Industrial Production and the Early Detection of urning Points, Università del Molise, Dipartimento di Scienze Economiche Gestionali e Sociali, Quaderni di Ricerca, Serie Statistica, n. 4. **Forthcoming in Empirical Economics**, 2003.

CHIU, C., MUMTAZ,H.,& PINTER, G. Forecasting with VAR models: fat tails and stochastic volatility. **Bank of England Bank of England working**, 2017.

D'AGOSTINO, D., GAMBETTI, L., & GIANNONE, D. Macroeconomic forecasting and structural change, **Journal of Applied Econometrics** 28, 82-101, 2013.

FEITOSA, M & TABAK, B.M. A. Forecasting industrial production in Brazil: Evidence from a wavelet approach. **Expert Systems with Applications**, 2010.

HASSANI, H. HERAVI, S & ZHIGLJAVSKY, A. Forecasting European industrial production with singular spectrum analysis, International. **International Journal of Forecasting**, 2009, 25: 103–118.

HERRIGEL, G. Globalization and the German industrial production model. Journal for Labour Market Research 48.2, 133-14, 2015

HOLLAUER, G., ISSLER, J., & NOTINI, H.H. Prevendo o crescimento da produção industrial usando um número limitado de combinações de previsões. Economia Aplicada, v.12, n. 2, p. 177 – 198, 2008.

LAPO, V. et al. Efficiency of investment stimulation methods in a timber industry complex: An econometric research. **Applied Econometrics**, v. 33, n. 1, p. 30-50, 2014.

MARKWALD, R. A., MOREIRA, A.R., & PEREIRA, P.L. Previsão da produção industrial: indicadores antecedentes e modelos de série temporal. **Pesquisa e Planejamento Econômico**, v. 19, n. 2, p. 233 – 254, ago, 1989.

PERRON, P. The great crash, the oil price shock, and the unit root hypothesis. Econometrica: **Journal of the Econometric Society**, p. 1361–1401, 1989.

PERRON, P. & VOGELSANG, T. J. Additional tests for a unit root allowing for a break in the trend function at an unknown time. **International Economic Review**, p. 1073–1100, 1998.

PRIMICERI, G. Time Varying Structural Vector Autoregressions and Monetary Policy, **Review of Economic Studies** 72, 821-852, 2005.

SIMPSON, P.W. OSBORNA, D. R & SENSIER, M. Forecasting UK industrial production over the business cycle. **International Journal of Forecasting**, 2001, 20: 405–424.

THURY, G., WITT, S. Forecasting Industrial Production Using Structural Time Series Models. **Omega**, Int. J. Mgmt Sci. Vol. 26, No. 6, 1998.

ZUANAZZI, P.T., & ZIEGELMANN, F. Previsões para o crescimento do PIB trimestral brasileiro com séries financeiras e econômicas mensais: uma aplicação de MIDAS. **Economia Aplicada**, v.18, n. 2, p. 295 – 318, 2014.

Appendix

Table 5 – Information criteria for the VAR models

Lag	AIC	SC	HQ
0	- 20.16799	- 19.56542	- 19.92312
1	-20.95986	- 20.01296*	- 20.57507*
2	- 20.89688	- 19.60565	- 20.37216
3	- 20.87078	- 19.23522	- 20.20613
4	- 20.85760	- 18.87770	- 20.05302
5	- 21.01692	- 18.69270	- 20.07242
6	- 21.05751	- 18.38896	- 19.97308
7	- 21.03081	- 18.01793	- 19.80646
8	- 21.080224*	- 17.72303	- 19.71596

Source: Elaborated by the author, based on the results obtained by the E*views software*.