International Journal of Statistics and Economics;
[Formerly known as the "Bulletin of Statistics & Economics" (ISSN 0973-7022)];
ISSN 0975-556X; Year: 2016, Volume: 17, Issue Number: 1; Int. j. stat. econ.;

Forecasting Brazilian Industrial Production Index with level and trend changes after crisis and SARIMA models

A. P. Alencar[1] and **F. M. M. da Rocha**[2]

[1]Institute of Mathematics and Statistics, University of São Paulo,
Rua do Matão 1010, São Paulo (Brazil);
Email : lane@ime.usp.br

[2]Institute, Federal University of São Paulo,
Rua Angélica, 100 Osasco (Brazil);
Email : f.marcelo.rocha@gmail.com

ABSTRACT

Two models are proposed for the monthly Brazilian industrial production index from 2002 to April 2015. The first model considers transient falls after the crisis in 2008 and another fall from December 2014 to April 2015, monthly fixed effects, different trends before and after 2008 and an autoregressive serial correlation for the errors. Another model without structural changes is the SARIMA model, including only a significant drop after the crisis in 2008. The first model identified a significant rising trend before the crisis of 2008, no significant growth after the crisis, significant declines just after the 2008 crisis and significant drop after December 2014. The SARIMA model do not assume different trends and levels but its forecasts indicates a worse scenario with no recovery in 2016.

Keywords: Time series, SARIMA, forecasting.

Mathematics Subject Classification: 62J12, 62G99

Computing Classification System: I.4

1. INTRODUCTION

In general, the industrial production index has been increasing in recent years, associated with the growth of the Brazilian economy. Recently some significant declines occurred in this index after the world crisis in 2008 and the deceleration of the Brazilian economy in 2015.

The crisis in 2008, known as subprime crisis that began in the United States, affected the economy of many countries which were affected by recession and are still recovering. In Brazil, industrial production shows a significant drop in late 2008 and in 2009, which partly stalled industrial activity.

Another significant drop of the industrial activity occurred in early 2015 related to declining growth expectations, successive interest rate increases and rising inflation. Another relevant factor is the shortage of electricity, which greatly increased energy prices, which directly affect some industries that rely heavily on electricity to function.

In this context, it is interesting to measure the effect of the 2008 crisis, assessing its impact in the short term, with level changes of the industrial production index, and the medium and long term due to changing trends in their behavior. Furthermore, it is also measured the effect of another crisis in 2015.

In order to measure these changes, a regression model with serial correlation is fitted for the monthly industrial production index from January 2002 to April 2015, and forecasts for the coming year were obtained, ie, from May 2015 to December 2016.

The main objective of this work is to propose a model that takes into account level changes of industrial production due to the crisis of 2008 and since the end of 2014 and trend changes since the crisis of 2008. On the other hand, the usual SARIMA model without several deterministic changes of level and trend, excepting a level change soon after the crisis in 2008, was proposed as a more conservative model. Forecasts and the corresponding confidence intervals using both models were compared to assess if the pessimist forecasts would be only a consequence of the choice of the level and trend changes.

2. Methodology

The monthly industrial production index analyzed in this paper is measured by the Brazilian Institute of Geography and Statistics (IBGE) based on data collected in the Monthly Industrial Research (IBGE,2004). This index is calculated using a new methodology since 2004 but IBGE provided the series since 2002 with average equal to 100 in 2012. In this paper we considered the industrial production index from 2002 to April, 2005. The methodology to calculate the index is detailed in IBGE (2004).

The first model considered is a regression linear model with changing level and trend, including:

- a trend component from 2002 to 2008 and a different trend since 2008;
- crisis08 variable equal to 1 from Dec, 2008 to May, 2009 and 0 otherwise;
- crisis08b equal to 1 from June to August, 2009 and 0 otherwise;
- y2015 equal to 1 from Dec, 2014 to April, 2015 and 0 otherwise;
- dummy variables I(Mt=i) equal to 1 if the t-th month is the i-th month of the year and 0 otherwise, for i= 1,…, 11, to model the seasonality with December as the reference month;
- AR(4) errors to model the serial correlation.

This first model for the industrial production index at time t is defined as

$$y_t = \beta_0 + \beta_1 (t-84) I(t<84) + \beta_2 (t-84) I(t \geq 84) + \beta_3 crisis08 + \beta_4 crisis08b + \beta_5 y2015 + \gamma_i I(M(t)=i) + \varepsilon_t ,$$

with t=1,…,160, i= 1,…, 11, and I(t<84) is a dummy variable equal to 1 if t<84 and 0 otherwise.

The parameter β_1 measures the linear trend until November, 2008, before the crisis in 2008, where the observation 84 corresponds to December, 2008. The linear trend after crisis is denoted by β_2 .Two level changes are included in this model with parameters β_3 and β_4, where the first measures the decrease just after the crisis and the second measures a posterior impact of this crisis. Another change level occurs after November 2014.

The error variables constitute a stationary stochastic process with zero mean and non-null autocorrelation. An autoregressive moving average (ARMA) model is proposed for these error variables after analysis of the correlogram and partial autocorrelation function of the residuals. Then, an autoregressive model of order 4 (AR(4)) model is proposed for these errors, as

$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \phi_3 \varepsilon_{t-3} + \phi_4 \varepsilon_{t-4} + w_t$, where w_t is a Gaussian white noise process, ie, a homoscedastic zero-mean process with null autocorrelations. The error variance is denoted by σ^2.

An alternative model (Model 2) is the SARIMA model that includes only the crisis08 variable as explanatory variable because it is necessary to include this variable to control the significant decrease just after the crisis of 2008. The choice of order of the SARIMA model was based on the Akaike criterion (AIC), which consists of maximizing the likelihood including a penalty to avoid a large number of parameters. The SARIMA(2,1,1)(1,1,2)12 with the exogenous variable crisis08 is specified as

$y_t = \beta_3 crisis08 + \varepsilon_t$,

$$\left(1 - \Phi_1 B^{12}\right)\left(1 - \phi_1 B - \phi_2 B^2\right)\Delta^{12}\Delta\varepsilon_t = \left(1 - \Theta_1 B^{12} - \Theta_2 B^{24}\right)\left(1 - \theta_1 B\right)w_t,$$

where Δ is the simple difference operator, Δ^{12} is the seasonal difference operator, B is the backshift operator and w_t is a Gaussian white noise with variance σ^2. The SARIMA model with exogenous variables are discussed for example in Shumway and Stoffer (2010).

The model was fitted using the conditional maximum likelihood method with the decomposition of the prediction error implemented in the forecast library for the R software (Hyndman, 2015). The auto.arima function allows to identify the order of the best SARIMA model minimizing the AIC (Hyndman and Athanasopoulos, 2013).

The residual analysis evaluated the assumptions of Gaussian white noise errors w_t, using qqplots to compare the residuals with the Gaussian quantiles and residual autocorrelation plots. The normality was also evaluated with the Shapiro-Wilk test (Shapiro and Wilk, 1965), since its power may be larger than other usual tests as compared by simulation in Razah and Wah (2011). The hypothesis of null autocorrelations until lag 12 and 24 was tested using the Ljung-Box test (Ljung and Box, 1978). The Breusch-Pagan test of homocedasticity of the errors w_t included all the exogenous variables in model 1 (Wooldridge, 2012).

The 5% significance level is adopted for the conclusion of all tests.

3. RESULTS

The industrial production index depicted in Figure 1 presents clear seasonal pattern, relevant growth from 2002 until 2008, a sharp decline in the end of 2008 until mid-2009 and it remains around 100 from 2010 to 2014. In the last 5 months, from Dec. 2014, the index presents a new decline.

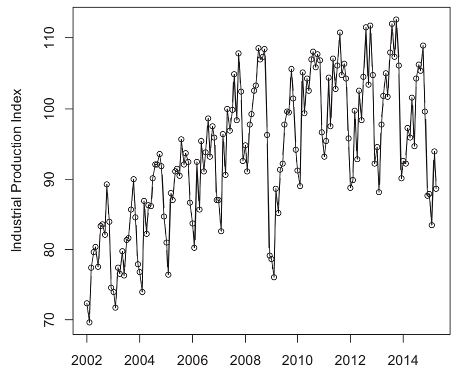

Figure 1: Monthly industrial production index

The regression model with the AR(4) errors presented significant autocorrelations only at lags 16 and 22, close to -0.2 as presented the sample autocorrelation function at Figure 2. The SARIMA residuals behave as a white noise with null autocorrelations. The Ljung-Box test confirmed that residuals present no correlation until lag 12 for both models (p=0.8024 and p= 0.337), but the correlations are not all null until lag 24 for the regression model with AR(4) errors (p=0.006) and are null for SARIMA model (p=0.119). Both residuals are distributed as Gaussian variables (p=0.954 and p=0.092 for models 1 and 2). Using the Breusch-Pagan test, the residuals seems homocesdatic (p=0.981 and p= 0.111 for models 1 and 2).

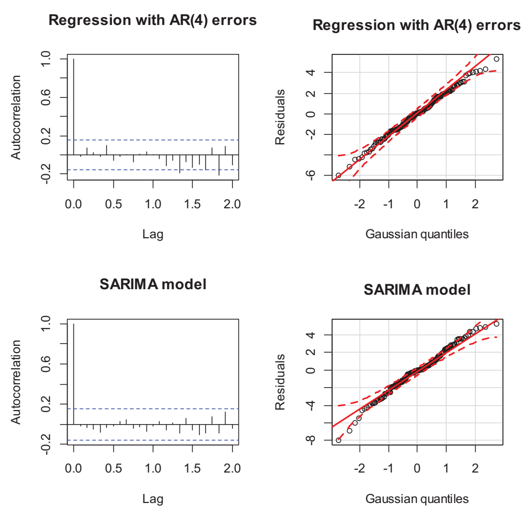

Figure 2: Diagnostic plots

The estimates presented at Table 1 indicate that the production index increased in average 0.3 each month until the 2008, but the trend was not significant after 2008 (p=0.383). In the period from

December, 2008 to May, 2009, the industrial production index presented a significant mean decrease of 13.3 ($p<0.001$) and from the June to August, 2009, the decrease was attenuated, but significant, corresponding to 7.0. These decreases occurred only in these periods and they were transitory, ie, the index returned to the level before the crisis, but it did not increased again because post-crisis trend was not significant. Another significant decreased occurred in December, 2014, making the index fall significantly 6.6 points since then.

With respect to the seasonality, February is the month with the smaller mean industrial production index and this expected production is significantly smaller than the observed in December. January presented a mean production index similar to the mean in December ($p=0.383$). All other months presented larger mean productions mainly from July to September.

The autoregressive term corresponding to the lag 2 presented no significant effect ($p=0.075$) and was removed from the final model.

Table 1: Estimates, standard error and p-values – Regression model with AR(4) errors

Parameter	Estimate	SE	P
Intercept	95.21	0.91	<0.001
trendtil08	0.30	0.01	<0.001
trendpost08	-0.01	0.02	0.383
M1	-1.00	0.83	0.231
M2	-3.72	0.91	<0.001
M3	6.30	0.75	<0.001
M4	3.25	1.02	0.001
M5	8.86	1.06	<0.001
M6	5.74	0.91	<0.001
M7	10.18	1.07	<0.001
M8	12.95	1.05	<0.001
M9	9.52	0.78	<0.001
M10	13.78	0.94	<0.001
M11	8.85	0.86	<0.001
Crisis08	-13.32	1.52	<0.001
Crisis08b	-6.98	1.51	<0.001
Year15	-6.58	1.59	<0.001
AR1	0.36	0.07	<0.001
AR3	0.42	0.07	<0.001
AR4	-0.34	0.08	<0.001
σ^2	4.255		

The SARIMA (2,1,1)(1,1,2)12 model was fitted to the industrial production index. This corresponds to a difference of the industrial production index and a seasonal difference to obtain a stationary series. After these differences, the series presented a relevant decrease from December, 2008 to May, 2009, what made necessary the inclusion of the effect of the crisis of 2008. The seasonal moving average term of order 1 was not significant ($p=0.533$) and was removed from the model.

Table 2: Estimates, standard error and p-values – SARIMA model

Parameter	Estimate	SE	P
AR1	-0.991	0.120	<0.001
AR2	-0.539	0.076	<0.001
MA1	0.512	0.129	<0.001
SAR1	-0.673	0.098	<0.001
SMA2	-0.609	0.120	<0.001
Crisis08	-8.005	1.516	<0.001
σ^2	6.998		

The estimated variance of the error is larger for the SARIMA model than for the regression model with AR(4) errors.

Figure 2 presents the industrial production index from 2012 and the forecasts from May, 2015 to December 2016 were obtained using both models. Until October, 2015 the forecasts are very similar, but since November, 2015, forecasts using the SARIMA model are smaller than the forecasts obtained using the regression model. It is worth to note that the confidence intervals with 95% level of confidence are larger for the SARIMA model due to the larger estimated error variance.

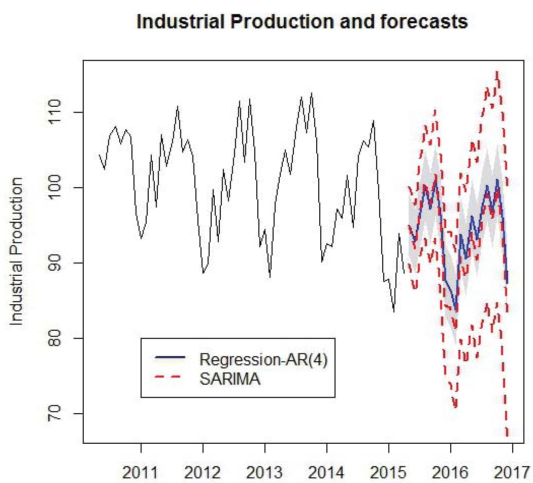

Figure 3: Industrial Production Index from 2012 and forecasts

Based on the SARIMA model, the industrial production index changes 3.78% in 2015, what corresponds to the percentage change forecast for December 2015 in relation to the index in December 2014. Another decrease of -1.1% is expected for 2016. Based on the regression model, the industrial production may vary only 0.1% in 2015 and -0.5% in 2016.

4. DISCUSSION AND CONCLUSION

The regression model allowed the identification of different trends before and after the crisis that occurred in 2008 and a significant transitory decrease after this crisis. These decreases corresponded to a larger one just after the crisis, identified since December, 2008 to May, 2009, and another one, with less impact, from June to August, 2009. Another significant decline occurred in 2015.

Both models do not presented large deviations from the assumptions of each model, but the regression model with the AR(4) errors presented some few residual autocorrelations close to 0.2 for lags 16 and 20. Even though we avoided to include more autoregressive or moving average terms, since including terms with order 16 or 20 would have no reasonable practical reason and this would imply a smaller number of observations to estimate the model.

It is worth noting that the error variance is larger for the SARIMA model, but this is a model for differences of a nonstationary series, with stochastic trends, not directly for the industrial production index. Moreover, the variance of forecasts are increasing over time (details in Hamilton, 1994) what is expected for stochastic trends and is very reasonable since for the long term, there is greater uncertainty. Model 1 do not have larger variability along time but is based on maintaining constant trends and levels.

The forecasts obtained using the regression model indicates that our industrial production index will continue to decrease. This fall must occur due to the non-significant trend after 2008 and the significant decline identified by the end of 2014. Therefore, these forecasts could be considered model-based with a pessimist bias including all level and trend decreases, but they constitute a better scenario than the SARIMA forecasts.

On the other hand, the SARIMA model seems to be a more unbiased model, called as a data-driven model, not including fixed linear trends and recent decline effects. Even though, the SARIMA forecasts also indicate a decline and since November, 2015 the forecasts are smaller than the obtained from the regression model. They also agree with the expectations in the Brazilian market, as experts believe that the industrial production index may fall 3.65% in 2015 as published in the Focus report (Brazilian Central Bank, 2015) and based on the SARIMA model the index may fall 3.8% in 2015. Therefore, it seems that our perspective is not very promising for 2016 and the current recession shall last long.

5. REFERENCES

Brazilian Central Bank, 2015. *Relatório Focus – June 2015*. Available at http://www.bcb.gov.br/pec/GCI/PORT/readout/R20150619.pdf.

IBGE, 2004. *Indicadores conjunturais da indústria - Produção Física*. Série Relatórios Metodológicos. Rio de Janeiro. Available in <http://www.ibge.gov.br/home/estatistica/indicadores /industria/pimpfbr/srmindconjind.pdf>.

Hamilton, J.D. , 1994. *Time series analysis*. Princeton University Press.

Hyndman, R.J. and Athanasopoulos G., 2013. *Forecasting: principles and practice*. Otexts.

Hyndman R.J., 2015. *forecast: Forecasting functions for time series and linear models*. R package version 6.1. Available at <http://github.com/robjhyndman/forecast>.

Ljung, G. M. and Box, G. E. P., 1978, On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

Razali, N. M. and Wah, Y. B., 2011, Power comparisons of Shapire-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson Darling tests. *Journal of Statistical Modeling and Analytics* **2**, 21-31.

Shapiro, S.S. and Wilk, M. B., 1965, An Analysis of Variance Test for Normality. *Biometrika* **52**, 591-611.

Shumway R.H. and Stoffer, D.S., 2011, *Time Series Analysis and Its Applications with R Examples*. 3rd Edition Springer.

Wooldridge, J. M., 2012, *Introductory Econometrics: A Modern Approach*. Cengage Learning; 5th edition.