Leonardo De Paoli Cardoso de Castro

# Forecasting Industrial Production in Brazil using many Predictors

## DISSERTAÇÃO DE MESTRADO

DEPARTAMENTO DE ECONOMIA
Programa de Pós-graduação em Macroeconomia e Finanças

Rio de Janeiro
May 2016

Pontifícia Universidade Católica
do Rio de Janeiro

**Leonardo De Paoli Cardoso de Castro**

# Forecasting Industrial Production in Brazil using many Predictors

**Dissertação de Mestrado**

Dissertation presented to the Programa de Pós–Graduação em Macroeconomia e Finanças of the Departamento de Economia, PUC-Rio as a partial fulfillment of the requirements for the degree of Mestre em Macroeconomia e Finanças.

Advisor: Prof. Marcelo Cunha Medeiros

Rio de Janeiro
May 2016

Pontifícia Universidade Católica
do Rio de Janeiro

**Leonardo De Paoli Cardoso de Castro**

# Forecasting Industrial Production in Brazil using many Predictors

Dissertation presented to the Programa de Pós–Graduação em Macroeconomia e Finanças of the Departamento de Economia, PUC-Rio as a partial fulfillment of the requirements for the degree of Mestre em Macroeconomia e Finanças. Approved by the following comission:

**Prof. Marcelo Cunha Medeiros**
Advisor
Departamento de Economia – PUC-Rio

**Prof. Carlos Viana de Carvalho**
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

**Prof. Eduardo Fonseca Mendes**
University of New South Wales – UNSW

**Prof. Monica Herz**
Coordinator of the Centro de Ciências Sociais – PUC-Rio

Rio de Janeiro, May 30th, 2016

**Leonardo De Paoli Cardoso de Castro**

Graduated in Economics from PUC-Rio in 2012.



Bibliographic data

de Castro, Leonardo De Paoli Cardoso

Forecasting Industrial Production in Brazil using many Predictors / Leonardo De Paoli Cardoso de Castro; Advisor: Marcelo Cunha Medeiros. – Rio de Janeiro : PUC-Rio, Departamento de Economia, 2016.

v., 42 f: il. ; 29,7 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui referências bibliográficas.

1. Economia – Tese. 2. Produção industrial. 3. Projeções. 4. LASSO. 5. Encolhimento. 6. Seleção de modelos. 7. Indicadores antecedentes. I. Medeiros, Marcelo Cunha. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 620.11

## Acknowledgment

To my advisor, Marcelo Cunha Medeiros, for his continued patience, guidance and advice.

To the members of my committee, professors Carlos Viana de Carvalho and Eduardo Fonseca Mendes, for their comments and suggestions.

To my friends (and mostly, mentors), Pedro Castro and Gabriel Hartung, for the lessons and experiences shared over the years.

To my mother, and my sister, for the kindness and support at all times. To Luiz and to Dudu, my brothers and best friends. To my two fathers, Fernando and Marco Antonio, which will always be my greatest examples of dedication and perseverance.

To Camila, who after 8 years finally decided to become part of my life.

## Abstract

de Castro, Leonardo De Paoli Cardoso; Medeiros, Marcelo Cunha (Advisor). **Forecasting Industrial Production in Brazil using many Predictors**. Rio de Janeiro, 2016. 42p. MSc. Dissertation – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

In this article we compared the forecasting accuracy of unrestricted and penalized regressions using many predictors for the Brazilian industrial production index. We focused on the least absolute shrinkage and selection operator (Lasso) and its extensions. We also proposed a combination between penalized regressions and a variable search algorithm (PVSA). Factor-based models were used as our benchmark specification. Our study produced three main findings. First, Lasso-based models over-performed the benchmark in short-term forecasts. Second, the PSVA over-performed the proposed benchmark, regardless of the horizon. Finally, the best predictive variables are consistently chosen by all methods considered. As expected, these variables are closely related to Brazilian industrial activity. Examples include vehicle production and cardboard production.



## Keywords

Industrial production; Forecasting; LASSO; Shrinkage; Model selection; Leading indicators;

## Resumo

de Castro, Leonardo De Paoli Cardoso; Medeiros, Marcelo Cunha (Orientador). **Prevendo a Produção Industrial Brasileira usando muitos Preditores**. Rio de Janeiro, 2016. 42p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Nesse artigo, utilizamos o índice de produção industrial brasileira para comparar a capacidade preditiva de regressões irrestritas e regressões sujeitas a penalidades usando muitos preditores. Focamos no *least absolute shrinkage and selection operator* (LASSO) e suas extensões. Propomos também uma combinação entre métodos de encolhimento e um algorítmo de seleção de variáveis (PVSA). A performance desses métodos foi comparada com a de um modelo de fatores. Nosso estudo apresenta três principais resultados. Em primeiro lugar, os modelos baseados no LASSO apresentaram performance superior a do modelo usado como *benchmark* em projeções de curto prazo. Segundo, o PSVA teve desempenho superior ao *benchmark* independente do horizonte de projeção. Finalmente, as variáveis com a maior capacidade preditiva foram consistentemente selecionadas pelos métodos considerados. Como esperado, essas variáveis são intimamente relacionadas à atividade industrial brasileira. Exemplos incluem a produção de veículos e a expedição de papelão.



## Palavras-chave

Produção industrial; Projeções; LASSO; Encolhimento; Seleção de modelos; Indicadores antecedentes;

# Contents

List of Abreviations **8**

1 Introduction **9**

2 Shrinkage regressions and the Lasso **13**
2.1 Penalized Regressions 13
2.2 Penalized Variable Selection Algorithm (PSVA) 16

3 Data and Period of Analysis **18**

4 Main Results **21**

5 Conclusion **25**

References **27**

A Estimation Procedures **30**

B Data **32**
B.1 Candidate Variables 32
B.2 Sources 36

C Figures **37**

## List of Abreviations

ADALASSO – Adaptive Least Absolute Shrinkage and Selection Operator
ADF – Augmented Dickey-Fuller
AR – Auto-Regressive
ARIMA – Auto-Regressive Integrated Moving Average
BIC – Bayesian Information Criteria
DFA – Dynamic Factor Analysis
DM – Diebold-Mariano
FGV - Fundação Getúlio Vargas
GDP – Gross Domestic Product
IBGE – Instituto Brasileiro de Geografia e Estatística
IPI – Industrial Production Index
LASSO – Least Absolute Shrinkage and Selection Operator
MA – Moving Average
MAE – Mean Absolute Error
MDIC – Ministério da Indústria, Comércio Exterior e Serviços
NTD – National Treasury Department
OLS – Ordinary Least Squares
PCA – Principal Component Analysis
PIM – Produção Industrial Mensal
PVSA – Penalized Variable Search Algorithm
RMSE – Root Mean Squared Error
VAR – Vector Auto regression

# 1
# Introduction

After GDP, the industrial production index (IPI) is the most important business cycle indicator for a given country. One of the main advantages of using the IPI in comparison to GDP is that the latter is usually disclosed in a quarterly manner while the former is released on a monthly basis. This gives us reliable high-frequency updates on the economic situation and helps market agents to assess current and future developments. Thus, it's safe to say that the IPI is a reliable indicator of the overall level of economic activity. In Brazil, the industrial sector accounts for over 17 percent of the economy's total GDP. Therefore, accurately forecasting short-term activity is essential for aiding the decision-making processes not only of industry-related companies but also of central bankers and other policy makers.

In this work, we addressed different methods for predicting both short- and long-term developments of the Brazilian industrial production index in a data-rich environment. While an extensive literature has focused on forecasting activity in real time for developed countries, few studies have been done for the Brazilian economy. Stock and Watson (2002) used factor-based models to forecast different real variables in the United States, including industrial production, over both short- and long-term horizons. Their results showed that high-dimensional forecasting methods over-performed simpler linear models in most cases. Bulligan, Golinelli and Parigi (2010) compared different forecasting methods for the Italian IPI series and found that factor models over-performed standard ARIMA regressions. Acedanski (2013) applied a similar methodology to the Polish industrial production index and concluded, contrary to Bulligan, Golinelli and Parigi's (2010) result, that simpler models often over-performed factor models. Lupi and Bruno (2001) used VAR models to forecast Italian industrial production and concluded that current leading indicators provide good predictive ability for not only short-but also long-term horizons. Apart from pursuing the best forecasting method, these authors also demonstrated the benefits of model combination following Stock and Watson (2006). Finally, Cunha (2010) used a diffusion index to anticipate future movements of the slack in Brazilian industrial production. His results show that factor models based solely on financial variables over-perform more complex models. This

result goes accords with those found by Hollauer, Issler and Notini (2008) and contradicts common wisdom, which states that hard-data variables are those most likely to present higher predictive power. Our main objective in this paper is to forecast Brazilian IPI's monthly growth by considering a large number of potential variables, using both soft and hard data.

Forecasting economic activity is not an easy task, especially when working with a large number of potential predictors. Since most candidates are mere approximations of reality, the uncertainty over the correct choice of regressors grows with the number of available candidates. Model specification is therefore an issue that is not readily resolved.

Standard variable selection algorithms usually rely on some type of information criteria, such as the Bayesian Information Criterion (BIC), a sequential testing procedure, such as forward stepwise selection, or some sort of model selection algorithm. The main problem with these approaches is that their ability to provide reliable forecasts is limited to small datasets. There are two reasons for this: first, the number of available models grows exponentially with the number of candidate variables, making these methods computationally costly or even unfeasible to evaluate all possible combinations; second, when the number of covariates is larger than the number of available observations ($k > T$), models become under-identified, making ordinary least squares estimation impossible. Thus, estimating high-dimensional models using standard procedures does not seem to be a good choice.

Empirical evidence helps us by providing priors on which leading economic indicators are most likely to anticipate the movements of other variables. In this work, our prior is that a small number of variables are relevant in determining future developments of the Brazilian Industrial Production Index (PIM). This implies the use of a sparse (small number of relevant variables) rather than dense (many relevant variables) dataset. The statistical method used in each of these two representations varies. Standard estimation methods for dense matrices are Principal Component Analysis (PCA), Dynamic Factor Analysis (DFA) and other variations. These methods capture the common variance between the explanatory variables and narrow these variables to a small number of factors. Despite providing good predictive results, they provide no information on which variables best describe the model and are also subject to bad quality data. Sparse matrix estimation, by contrast, is generally achieved using penalized regressions. The introduction of a penalty parameter into standard linear regressions helps to diminish model complexity by forcing the coefficients of irrelevant variables toward zero. The Ridge Regression proposed by Hoerl and Kennard (1970) and the Elastic Net regression proposed

by Zou and Hastie (2005) are examples of penalized least squares.

Statistical data on the Brazilian economy have only recently begun to be gathered. Therefore, forecasting the country's economic variables can be problematic since we face a "fat" shaped dataset (many variables but not as many observations, $k > T$). Model selection in this type of dataset is subject to various issues, e.g., over-fitting and constant revisions. Our main goal in this paper is to forecast the Brazilian industrial production index by considering a large number of potential explanatory variables. We use three different approaches to cope with this problem. First, we use factor-based models as our benchmark. Second, we use shrinkage regressions (or penalized regressions), using the Lasso (Least Absolute Shrinkage and Selection Operator) proposed by Tibshirani (1994) and the Adalasso (Adaptive Least Absolute Shrinkage and Selection Operator) proposed by Zou (2006). Shrinkage regressions are designed to select the most parsimonious model while preserving the true representation of data generating process of the response variable. Third, we combine penalized regressions with a variable selection algorithm that we will henceforth refer to as PVSA (Penalized Variable Search Algorithm). This methodology reduces the number of potential explanatory variables using shrinkage regressions and then uses a variable search algorithm to select the model with the best out-of-sample performance. Finally, we also assess the predictive ability of the simple mean of all forecasts produced by the PSVA. We will henceforth refer to this forecast combination as POOL.

Our results demonstrate that shrinkage regressions over-perform our benchmark in short-term forecasts. The combination of penalized regressions with a variable selection model (PVSA) over-performed the Lasso, the Adalasso and the proposed benchmark at all forecasting horizons. Furthermore, pooling forecasts using the Lasso as a first-step variable filter (POOL) achieved the best forecasting accuracy in our exercise. Diebold-Mariano tests of statistical equivalence corroborate this outcome. Moreover, the shrinkage models predictive ability and variable selection consistency worsened as we increased the number of candidates, revealing a methodological limitation. The results also show that long-term forecasts exhibit inferior performance compared with shorter horizons and are often beaten by the response variable's historical mean. A possible explanation for this result is that the available leading indicators seem to contain little information about long-run investment decisions and, consequently, future production.

Finally, although our main objective is to select the most accurate model independent of the forecasting horizon, all methods yielded similar sets of candidate variables. As we would expect, those most frequently selected

were related to economic activity and production. Individually, explanatory variables related to the auto industry were by far the most common in our analysis, suggesting that Brazilian industrial production is highly dependent on this segment. Another interesting result is that business sentiment surveys also exhibited high predictive power and were frequently selected in both short- and long-term forecasts. This corroborates the results observed in Hollauer, Issler and Notini (2008) and Mello, E. (2014) that show that the industry's capacity utilization index and business sentiment surveys significantly enhance forecasting power.

This paper is organized as follows. Section 2 introduces the penalized regression framework, focusing on the Lasso and its variants. Section 2 also presents the PSVA. In Section 3, we briefly discuss the Brazilian IPI and the candidate variables available in our dataset. In Section 4, we present our main results, and we conclude our analysis in Section 5.

# 2
# Shrinkage regressions and the Lasso

Approximating complex inter-correlations between real-world variables using small, semi-structural models is a difficult task. For instance, an economy's transmission mechanisms are determined by a huge number of different variables - which may or may not be directly related to the economy itself. Given the underlying uncertainty, accurate variable selection plays a substantial role in the pursuit of a single model that can simplify reality using a small number of candidates. Sequential testing procedures such as forward stepwise selection and complete subset regression analysis (Elliot, Gargano and Timmermann, 2012) achieve good results when working with small datasets by providing a parsimonious model with good predictive performance. However, these methods become inefficient and, eventually, unfeasible when using a large number of predictors, as when the number of covariates exceeds the number of observations ($k > T$), models become under-identified. Penalized regressions such as the Lasso are a fast and reliable alternative for high-dimensional estimation and forecasting.

In this chapter, we introduce the penalized regression framework. We also present a variant of penalized regression called the PSVA (Penalized Variable Selection Algorithm), which combines shrinkage methods and a variable selection algorithm.



## 2.1
## Penalized Regressions

Shrinkage regressions such as the Lasso can be viewed as a penalized functional form of standard ordinary least squares estimation. There are two main objectives motivating the introduction of tuning parameters into OLS regressions: (i) reducing model complexity by shrinking the estimates of the regression coefficients toward zero relative to their OLS estimates and (ii) improving accuracy by trading bias to reduce estimator variance. By producing some coefficients that are exactly zero, the Lasso provides interpretable models and is therefore an efficient alternative when working with high-dimensional datasets. The Lasso estimator is defined as follows:

$$\hat{\boldsymbol{b}} = \arg\min_{\hat{b}} ||\boldsymbol{Y} - \boldsymbol{X}b||_2^2 + \lambda \sum_{p=1}^{P} |b_p|, \tag{2-1}$$

where $\boldsymbol{b}$ is the $N \times 1$ vector of parameters, $\boldsymbol{Y} = (y_1, ..., y_t)'$ is the response variable, $\boldsymbol{X}$ is the $P \times n$ data matrix, and $\lambda$ is a non-negative regularization parameter. In penalized regressions, model complexity becomes a function of its regularization parameter ($\lambda$). As the tuning parameter increases in size, penalties become more conservative and reduce the model specification by shrinking irrelevant variables coefficients toward zero ($\lambda = \infty$ makes all coefficients equal to zero). Figure 2 illustrates this relationship by showing the sensitivity between the Lasso's penalization parameter (x axis) and the candidate regressors coefficients (y axis). Most studies select the tuning parameter using cross-validation[1] or some information criteria such as the Bayesian Information Criterion (BIC). The former is not always the best choice since it is not robust to inter-temporal dependence. In other words, cross-validation does not provide the expected results when the data are not independent and identically distributed (i.i.d.). Thus, regularization parameters selected through information criteria appear to work best for time-series datasets.

Although capable of efficiently handling large datasets, the Lasso also faces limitations. Zou (2006) and Meinshausen and Bühlmann (2006) showed that there are scenarios in which the model selection determined by the Lasso is inconsistent and does not achieve the oracle property (Fan and Li, 2001). In other words, the Lasso is not always capable of selecting the correct set of relevant variables, and even if it is, the Lasso estimator will most likely present a different asymptotic distribution than that provided by the OLS estimator (oracle). Zou (2006) suggested a two-step procedure called the adaptive Lasso (Adalasso), whereby individual weights are used to further constrain variable coefficients in the $l_1$ penalty. This modification allows the Adalasso to achieve variable selection consistency. Moreover, Medeiros and Mendes (2015) derived conditions under which the Adalasso presents both model selection consistency and enjoys the oracle property. The Adalasso estimator is defined as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\beta}} ||\boldsymbol{Y} - \boldsymbol{X}\beta||_2^2 + \lambda \sum_{p=1}^{P} w_p |\beta_p|, \tag{2-2}$$

where $\boldsymbol{\beta}$ is the $N \times 1$ vector of parameters, $\lambda$ is the first-step Lasso regularization parameter, $\boldsymbol{w}$ is a weight vector defined as $w_p = |\hat{\beta}_{Lasso}|^{-\gamma}$, with

[1] A model validation technique.

$\beta_{Lasso}$ as a first-step Lasso estimator to $\beta$, and $\gamma$ is a non-negative weighting parameter ($\gamma > 0$).

The main challenge in applying this approach is to correctly tune the adaptive Lasso or, in other words, to retrieve the optimal pair of parameters $(\gamma, \lambda)$ that will yield the model with the closest representation to the true data generation process. Both parameters are often selected through the minimization of some sort of loss function, e.g., root mean squared error or mean absolute error, using cross-validation methodology. As mentioned above, this method does not perform well in frameworks with serially correlated data. To address this issue, practitioners have resorted to the use of out-of-sample evaluation or information criterion, such as the BIC. Zhang et al (2010), Wang et al (2007) and Zou et al (2007) showed that the BIC is a reliable alternative to cross-validation in time-series frameworks. Finally, Medeiros and Mendes (2015) used the BIC as the selection method for both the first-step Lasso regularization parameter ($\lambda$) and the Adalasso weighting parameter ($\gamma$). They showed that this flexible version of the Adalasso is asymptotically consistent and enjoys the oracle property (selects the correct variable subset) even when the number of candidate variables is much larger than the number of observations.

In addition to the Lasso, there are also other types of penalized regressions that are worth highlighting. The Lasso is a particular case in which a standard ordinary least squares estimator is constrained by a linear absolute value penalty term. Other shrinkage methods such as the Ridge Regression and the Elastic Net also use quadratic penalties. Let us consider a generic loss function ($l$) that contains both linear ($l_1$) and quadratic penalties ($l_2$). The latter can be defined as follows:

$$l_{1,2} = \lambda \sum_{p=1}^{P} [(1-\alpha)|b_p| + \alpha |b_p|^2], \tag{2-3}$$

where $\lambda$ is a positive regularization parameter, $b_p$ represents the $p^{th}$ independent variable coefficient, and $\alpha$ represents the share of each penalty ($l_1$ or $l_2$) in the minimization procedure. The Lasso is a particular case in which $\alpha = 0$. As $\alpha$ increases toward *unity*, the quadratic penalty gains importance and may not necessarily shrink irrelevant variables to zero but, instead, to a very small value. In contrast to the Lasso, the Ridge Regression (the particular case in which $\alpha = 1$) does not produce a parsimonious model, as it always retains all the predictors in the model. When $\alpha$ is in between these particular cases, the following regression is known as the Elastic Net (Zou and Hastie, 2005), which uses a combination of both linear ($l_1$) and quadratic ($l_2$)

penalties. By combining the effects of both the Lasso and the Ridge Regression, the Elastic Net produces a sparse model with good predictive accuracy while encouraging a grouping effect (it selects highly correlated regressors).

## 2.2 Penalized Variable Selection Algorithm (PSVA)

Testing all possible variable combinations in high-dimensional datasets is a costly and sometimes unfeasible task. We propose a variation of the Complete Subset Regression method proposed by Elliot, Gargano and Timmermann (2013) by combining shrinkage regressions and a variable selection algorithm. We also propose a set of constraints that reduce the required computational effort and makes variable selection more efficient.

The methodology applied consists of two main parts. First, we reduce the number of available candidates to a few targeted predictors[2] by applying the Lasso as a first-step variable filter. Since our procedure involves re-estimating the model for each out-of-sample prediction (in our work, we consider a 36-step forecast sample), the Lasso may ultimately select heterogeneous models. The variables selected through the procedure compose a new set of candidates that is used in the second part of our proposed algorithm. By employing a lower-dimension dataset we are able to address conventional model selection methods such as sequential testing procedures and information criterion.

The second part of the algorithm uses a slightly different approach from ordinary variable selection methods by imposing two additional constraints. The first further reduces the set of candidates selected by the Lasso by imposing a cap ($k$) on the number of variables available to our algorithm. Second, instead of estimating all available variable combinations ($2^k$), we also limit the maximum number of variables that can be included in each model ($p$). Both modifications make our algorithm more efficient by eliminating from the analysis variables that may have been incorrectly selected by the Lasso and precluding model over fitting. Our results consider $k = 20$ and $p = 6$.

The top-$k$ most selected variables are considered the most likely to best represent the response variable's data generating process and are therefore included as candidates by our algorithm. If a number of variables greater than $k$ present the same (and highest) selection rate, all will be assigned as candidates. For example, assume that $k$=10 and that the Lasso selected 12 of the total available variables in its estimation procedure with the same selection rate. In this case, $k$ will be reset to 12, thus considering all selected variables as possible candidates.

[2]See Bai & Ng (2008) for more on targeted predictors.

Having selected $n \leq k$ variables, we estimate every possible variable combination that respects the $p$-maximum variable constraint using ordinary least squares. The total number of different model specifications is then given by the following equation:

$$tot = \sum_{i=1}^{p} \frac{n!}{(n-i)!i!}, \tag{2-4}$$

where $n$ is the number of available candidates, and $i$ is the number of variables combined in each individual model.

Finally, each individual model produces one-step-ahead forecasts within a moving window of $j$ periods. This procedure is repeated $Z$ times, and in each iteration, the moving window shifts backward by one period. To illustrate, suppose that in the first iteration, models are tested for their performance in the window between $t_{1,1} = v$ and $t_{2,1} = v + z$. The best model in this iteration, denoted $m_1$, will be the model with the best out-of-sample RMSE error within this window. Then, in the second iteration, the moving window would shift backward by one. As a result, models would be tested for their out-of-sample RMSE performance in the window between $t_{1,2} = v - 1$ and $t_{2,2} = v + z - 1$. We would then have another model, denoted $m_2$, which performed the best in this second window. Notably, $m_1$ and $m_2$ may or may not be the same. After $z$ iterations, we would have $Z$ best models, each of which had the top performance in its respective moving window. These models are then used to produce the "true performance" of our proposed algorithm. Considering the example above, $m_1$ would give us the forecast for period $v + z + 1$, while $m_2$ would produce the forecast for period $v + z$. The idea is to always predict using the model that had the best out-of-sample performance up to the last observation of a given moving window. Within each moving window, we also consider not only individual models but also model averages, such that $m_1$ forecasts may be an average over the *tot* models rather than one single model.

# 3
# Data and Period of Analysis

The PIM is Brazil's monthly industrial production index (IPI) and represents the product of both manufacturing and extractive industries. The index is calculated by the Brazilian Institute of Geography and Statistics (IBGE) and was first compiled in the 1970s. In May 2014, the IPI received its second methodological revision, updating the industries' activities and their index weights to the latest classification of the National Classification of Economic Activities (CNAE 2.0). Despite having changed its methodology in 2014, the IBGE updated the IPI in January 2012 and linked it to its previous version, creating a series that begins in 2002.

Since the available data were subject to a methodological revision, it is very unlikely that the IPI's data generating process has remained the same over time. Therefore, the choice of which sample to use in our analysis was not trivial. We had three possible options:

(1) Use only the new version (with the methodological update) of the IPI, running from January 2012 until September 2016. This would give us a sample with 48 observations.

(2) Use the old version (without the methodological update) of the IPI, running from January 2002 until September 2016. This would give us a sample with 120 observations.

(3) Test whether the coefficients suffered significant changes after the methodological update. If the test result were negative, we could estimate our models using the entire available sample, running from January 2002 until September 2016. This would give us 177 observations.

As estimating models with a small number of observations and a large number of candidates is a problem that lacks a simple solution, we chose the option that provided us with the largest sample available (option 3). To assess whether there was a significant change between periods, we estimated a regression with seasonal dummies and an additional control variable that represented the updated series period[1]. The results indicated that we could not

[1] A dummy equal to 0 from January 2002 to December 2011 and equal to 1 from January 2012 to September 2016.

significantly reject that the two periods presented the same seasonal effects. In light of this result, option 3 seemed to be the best choice for our analysis.

There is an additional consideration that needs to be made: our main objective is to perform real-time forecasts of the IPI's monthly growth. This means attempting to replicate the exact conditions (with the same information set) faced by forecasters or, in other words, considering only the preliminary nature of the data and discarding possible revisions. Therefore, our response variable was constructed by linking the non-revised realizations of the IPI's monthly growth from January 2002 until our last available observation in September 2016. Figure 3.1 compares the original (non-revised) and the revised industrial production index using January 2002 as our initial point[2]. Figure 3.2 shows the relative frequency of revisions made to the IPI over the years.

Following the methodology applied to construct our real-time IPI series, we selected over 70 candidate variables, of which we also include two lags. The lagged response variable is also included as a candidate variable in our dataset. All variables were tested for the presence of a unit root and transformed (first differentiated) when necessary. To account for the seasonality of the series, monthly dummies were also included as candidate regressors. The explanatory variables are both soft and hard data and are detailed in Appendix B. Our dataset contains variables that cover the manufacturing of intermediary, durable and non-durable goods, government debt, taxes, foreign trade, wages, unemployment, confidence indexes, sales, leverage, interest rates, etc. All candidate variables possess the same time span as the IPI (January 2002 to September 2016).

Our sources are Bloomberg, the Brazilian Institute of Geography and Statistics (IBGE), the Getúlio Vargas Foundation (FGV), the National Treasury department, the Brazilian Central Bank, the Ministry of Development, Industry and Foreign Trade, the Ministry of Labor, and the Industry Federation of the state of São Paulo (FIESP).

[2]January 2002 = 100.

Figure 3.1: IPI - Original and Revised Series. January 2002 = 100

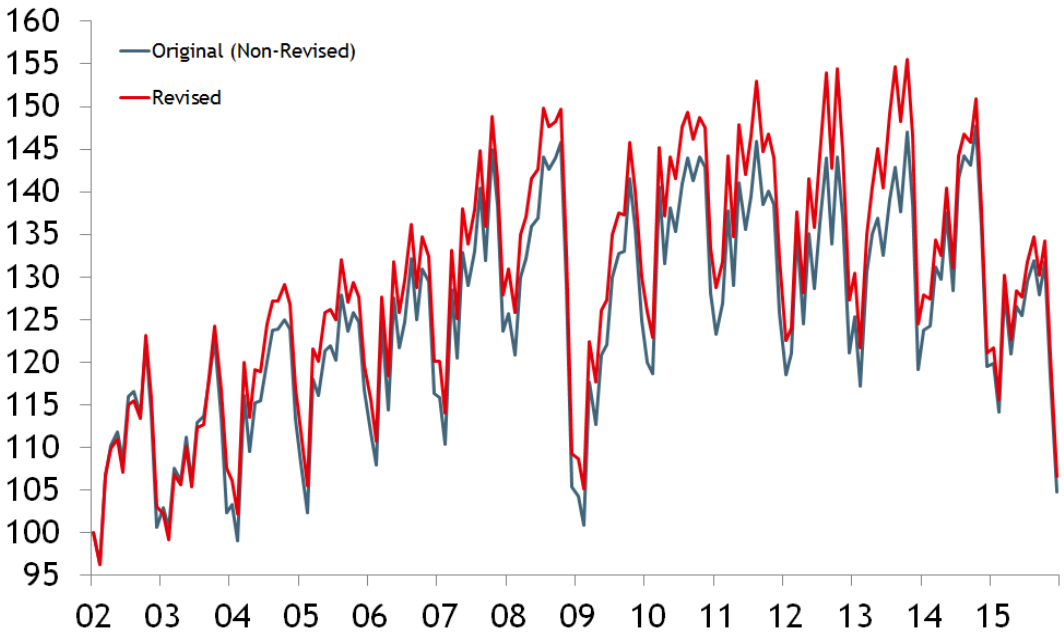
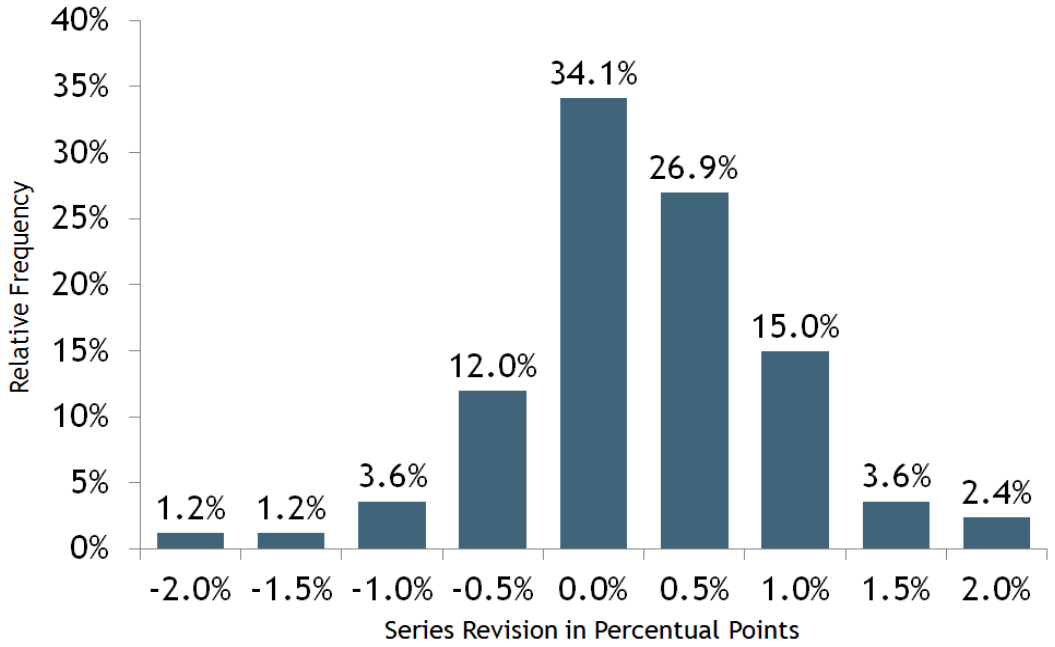





Figure 3.2: IPI's Revisions Relative Frequency



# 4
# Main Results

This section presents the performance of three forecasting methods used to predict the monthly growth of the Brazilian industrial production index (IPI) ($\frac{PIM_t}{PIM_{t-1}} - 1$). The first method is based on principal components analysis (PCA). The second method selects models through penalized regressions (shrinkage) using the Lasso and the Adalasso. The third and final method combines penalized regressions and a variable selection algorithm (PSVA). Finally, we also assess the predictive ability of the simple mean of all forecasts produced by the PSVA. We will henceforth refer to this forecast combination as POOL.

Our dataset is composed of over 70 candidates with 177 monthly observations each, dated from January 2012 to September 2016. We use up to two lags of the response and candidate variables as possible predictors. To account for the series seasonality, monthly dummies were also considered as candidates. All variables were tested for the presence of a unit root and transformed (first differentiated) when necessary. The models were estimated using a 36-month rolling window, and their forecasts accuracy were compared using the root mean squared error (RMSE) and the mean absolute error (MAE). All estimated models and predictive results are for up to 12-month-ahead horizons. The estimated equation is defined as follows:

$$Y_{t+h} = \alpha_0 + \sum_{i=0}^{2} \delta' X_{t-i} + \sum_{s} \theta_s D_s + \mu_{t+h}, \tag{4-1}$$

where $Y_{t+h}$ is the Brazilian industrial production index monthly growth, $\alpha_0$ is a constant term, $X_{t-i}$ is the vector containing the candidate variables and their lags, $D_s$ is a vector containing monthly seasonal dummies, and $\mu_t$ is an error term. This equation states that the IPI is highly dependent on calendar effects and past realizations of both the response variable and other leading indicators.

The first interesting result is that while the IPI's monthly growth was highly dependent on calendar effects, its past realizations lacked predictive power relative to the remaining predictors in our dataset. The seasonal component, by contrast, was very important for predicting movements in the

responses variable. Figure 3 shows that the series monthly results behave in a similar way over time. Moreover, when tested formally, the results indicated that the IPI exhibits strong seasonality. Figure 4 shows how the seasonal factors are spread across the year. This result led us to include monthly dummies as candidate regressors to address this seasonal behavior. Note that imposing dummies as candidates while using shrinkage methods can be problematic. For instance, it is unlikely that they will always be selected by the algorithm since other variables can capture part of the seasonality, thus causing the dummies to be regarded as irrelevant. Nevertheless, results have shown that both the Lasso and the Adalasso recurrently considered the seasonal dummies as relevant regressors.

By providing good forecasting results in the presence of a very large number of variables, penalized regressions out-perform factor-based models when exposed to a large number of irrelevant variables. For comparison purposes, we tested the Lasso, Adalasso and the PSVA using three different datasets:

(1) Dataset composed of the most relevant variables selected following our own prior. Total of 59 candidates.

(2) Dataset 1 plus 20 candidates (includes lagged variables). Total of 79 candidates.

(3) Dataset 1 plus 60 candidates (includes lagged variables). Total of 119 candidates.

The result of this exercise was that both the Lasso and the Adalasso became more conservative as we included a larger number of irrelevant variables as potential candidates. This becomes more clear in longer-term forecasts, in which only a small number of variables are considered relevant. By becoming more conservative, the forecasting performance at any given horizon deteriorated as the size of the dataset increased. Appendix B presents the variables used in each of the three previously mentioned groups. Figure 5 compares the methods one-step-ahead forecasting performance using each dataset. Figures 6, 7 and 8 depict the average number of selected variables at each forecasting horizon using the Lasso and the Adalasso, each with a different set of variables.

Using dataset number 2 as the available group of variables, the proposed methods achieved strong results in short-term forecasting, which is the most relevant horizon for econometric models. This outcome highlights the potential benefits of shrinkage regressions in data-rich environments. When compared to the one-step ahead forecasts disclosed by other market participants, the PSVA

and POOL were included among the top-10 best-performing predictors. This is another strong result, especially as market participants are able to introduce priors into their forecasts to cope with movements that are not captured by the available regressors. Figure 9 shows the top-ranked market participants forecasting performance and the ranking of the Lasso, the Adalasso, the PVSA and the POOL model. Furthermore, another interesting result was that the POOL's forecasting performance improved as the number of irrelevant variables grew. This leads us to two conclusions: first, shrinkage regressions were unable to efficiently discard all irrelevant variables; second, variables that were presumed to be orthogonal to the IPI captured part of the movement that other candidates were not able to capture.

The results also indicate that long-term forecasts exhibit considerably worse performance than those at shorter horizons and are often out-performed by the response variable's historical mean. A possible explanation for this result is that the available leading indicators seem to contain little information about long-run investment decisions and, consequently, future production. Figures 10 and 11 show the root mean squared error (RMSE) and the mean squared (MAE) at different forecasting horizons using the Lasso, the Adalasso, the PSVA, the POOL and a model using principal components (our benchmark).

It's worth noting that all of the proposed methods were efficient in choosing variables consistent with economic theory, regardless of forecasting horizon. Examples of recurrently selected variables in long-term forecasts were lagged interest rates and spread rates. A possible transmission mechanism regarding these variables is that long-term investments are negatively influenced by a cost effect, i.e., higher interest rates. At short-term horizons, the best-performing models regularly selected variables related to vehicle production, highway traffic, cardboard sales and working days. This result suggests that Brazilian industrial production is highly dependent on the country's auto industry.

To assess the statistical relevance of the various forecasts generated by the proposed methods, we used the Diebold-Mariano test for predictive equality. All forecasts were compared against our benchmark. These tests revealed that one-step-ahead predictions were statistically superior to the proposed benchmark's performance. Tests of the longer-term forecasts ($h > 1$), by contrast, did not reject the null hypothesis of statistical equality. This result persisted even after adjusting the DM test to improve small-sample properties, as proposed by Harvey, Leybourne, and Newbold (1997). Table 2 illustrates this analysis by presenting the adjusted DM test's p-values for 1- to 12-month-ahead forecasts.

Table 4.1: Adjusted Diebold-Mariano One-sided Tests: P-values (x 100)

| **Forecasting Horizon** | **Lasso** | **Adalasso** | **PSVA** | **Pool** |
|---|---|---|---|---|
| $t = 1$ | 4.08 | 8.85 | 0.44 | 0.47 |
| $t = 2$ | 16.21 | 16.28 | 3.06 | 5.02 |
| $t = 3$ | 16.41 | 16.26 | 4.78 | 5.61 |
| $t = 4$ | 17.20 | 17.78 | 7.23 | 6.92 |
| $t = 5$ | 19.04 | 19.03 | 0.61 | 1.00 |
| $t = 6$ | 20.60 | 19.71 | 2.46 | 3.75 |
| $t = 7$ | 15.60 | 9.18 | 0.17 | 0.07 |
| $t = 8$ | 23.19 | 23.08 | 9.83 | 9.93 |
| $t = 9$ | 23.40 | 23.20 | 2.11 | 5.94 |
| $t = 10$ | 28.31 | 23.73 | 7.68 | 8.26 |
| $t = 11$ | 19.98 | 19.63 | 6.79 | 6.79 |
| $t = 12$ | 20.79 | 23.18 | 8.29 | 12.89 |

# 5
# Conclusion

Forecasting economic variables in rich, high-dimensional environments is a difficult task. In this paper, we address this issue and compare different forecasting methods when applied to Brazilian data, specifically the industrial production index (IPI). Since standard linear models tend to become inefficient as our dataset enlarges, we utilized shrinkage methods to address the dimensionality problem. Here, we consider the Lasso and the Adalasso. We also propose a combination of penalized regressions and a variable search algorithm, which we refer to as PVSA. Finally, we also assess the predictive ability of the simple mean of all forecasts produced by the PSVA, which we call POOL. The forecasts produced by these previously mentioned methods were compared to a principal component model, our benchmark.

The results demonstrated that shrinkage methods over-performed the proposed benchmark in short-term forecasts, regardless of the number of available candidates. Furthermore, the PSVA and the POOL excelled the greater part of market participants predictions. This is a strong result given that individuals are able to introduce their priors to cope with movements that are not captured by the available regressors. Long-run forecasts, by contrast, exhibit considerably worse performance and often under-perform the response variable's historical mean.

Finally, all methods considered selected similar sets of explanatory variables. Short-term predictions were usually dominated by the auto industry related variables, highway traffic, cardboard sales, business days, and so forth. Long-term forecasting, by contrast, exhibited greater variability and selected variables such as interest rates, spread rates, confidence indexes and so forth. Moreover, the shrinkage models' predictive ability and variable selection consistency worsened as we increased the number of candidates, revealing a clear limitation of the minimization algorithm. The merger between penalized regressions and a deterministic variable selection algorithm addressed this problem by consistently selecting the same variable subset regardless of the number of available regressors. The short-term forecast results suggest that Brazil's industrial production is highly dependent on the country's auto industry and other related sectors. Medium-term determinants, by contrast, were unclear,

suggesting both methodological and data limitations.

## References


[1] ACEDÁNSKI, J.. **Forecasting industrial production in poland – a comparison of different methods**. Ekonometria Econometrics, 39:40–51, 2013.

[2] ARTHUR E. HOERL, R. W. K.. **Ridge regression: Biased estimation for nonorthogonal problems**. Technometrics, 12(1):55–67, 1970.

[3] BAI, J.; NG, S.. **Forecasting economic time series using targeted predictors**. Journal of Econometrics, 146(2):304–317, October 2008.

[4] BATES, M. J.; GRANGER, J. C. W.. **The combination of forecasts**. Journal of the Operational Research Society, 20(4):451–468, 1969.

[5] BRUNO, G.; LUPI, C.. **Forecasting euro-area industrial production using (mostly) business surveys data**. ISAE Working Papers 33, ISTAT - Italian National Institute of Statistics - (Rome, ITALY), 2003.

[6] BULLIGAN, G.; GOLINELLI, R. ; PARIGI, G.. **Forecasting monthly industrial production in real-time: from single equations to factor-based models**. Empirical Economics, 39(2):303–336, 2010.

[7] CUNHA, F. C. S.. **Previsão da Produção Industrial do Brasil: Uma Aplicação do Modelo de Índice de Difusão Linear**. Dissertação de mestrado, Departamento de Engenharia Elétrica da PUC-Rio, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

[8] DE MELLO, E. P. G.; FIGUEIREDO, F. M. R.. **Assessing the Short-term Forecasting Power of Confidence Indices**. Working Papers Series 371, Central Bank of Brazil, Research Department, Dec. 2014.

[9] ELLIOTT, G.; GARGANO, A. ; TIMMERMANN, A.. **Complete subset regressions**. Journal of Econometrics, 177(2):357–373, 2013.

[10] FAN, J.; R., L.. **Variable selection via nonconcave penalized likelihood and its oracle properties**. Journal of the American Statistical Association, 96:1348–1360, 2001.

[11] FRANCIS X. DIEBOLD, R. S. M.. **Comparing predictive accuracy**. Journal of Business & Economic Statistics, 13(3):253–263, 1995.

[12] HARVEY, D.; LEYBOURNE, S. ; NEWBOLD, P.. **Testing the equality of prediction mean squared errors**. International Journal of Forecasting, 13(2):281–291, 1997.

[13] HOLLAUER, G.; ISSLER, J. A. V. ; NOTINI, H. H.. **Prevendo o crescimento da produção industrial usando um número limitado de combinações de previsões**. Economia Aplicada, 12:177 – 198, 2008.

[14] MEDEIROS, M. C.; MENDES, E. F.. **l1-Regularization of High-Dimensional Time-Series Models with Flexible Innovations**. Journal of Econometrics, 191:255 – 271, 2016.

[15] MEINSHAUSEN, N.; BÜHLMANN, P.. **High dimensional graphs and variable selection with the lasso**. ANNALS OF STATISTICS, 34(3):1436–1462, 2006.

[16] STOCK, J. H.; WATSON, M.. **Forecasting with many predictors**. volumen 1, chapter 10, p. 515–554. Elsevier, 1 edition, 2006.

[17] STOCK, J.; WATSON, M.. **Macroeconomic forecasting using diffusion indexes**. Journal of Business and Economic Statistics, 20(2):147–162, 2002.

[18] TIBSHIRANI, R.. **Regression shrinkage and selection via the lasso**. Journal of the Royal Statistical Society, Series B, 58:267–288, 1994.

[19] WANG, H.; LI, G. ; TSAI, C.-L.. **Regression coefficient and autoregressive order shrinkage and selection via the lasso**. Journal of the Royal Statistical Society Series B, 69(1):63–78, 2007.

[20] ZHANG, Y.; LI, R. ; TSAI, C.-L.. **Regularization parameter selections via generalized information criterion**. Journal of the American Statistical Association, 105(489):312–323, 2010.

[21] ZOU, H.. **The adaptive lasso and its oracle properties**. Journal of the American Statistical Association, 101(476):1418–1429, 2006.

[22] ZOU, H.; HASTIE, T.. **Regularization and variable selection via the elastic net**. Journal of the Royal Statistical Society, Series B, 67:301–320, 2005.

[23] ZOU, H.; HASTIE, T. ; TIBSHIRANI, R.. **On the "degrees of freedom" of the lasso**. Ann. Statist., 35(5):2173–2192, 10 2007.

# A
# Estimation Procedures

The estimation procedure of the Lasso and its extensions was implemented using the GLMNET package contained in the R software repository. The default procedure used to select the penalization parameter ($\lambda_1$) is cross-validation. Results have shown that this procedure may produce misleading results in serially correlated datasets. To address this issue, we followed the methodology used by Medeiros and Mendes (2015). First, we estimated the regularization path using the Lasso. We then split this path into 100 consecutive parts, ranging from $\lambda = 0.1$ (low penalization) to $\lambda = 1$ (high penalization). For each $\lambda$ we had a different set of variables and therefore a different set of in-sample estimates for our response variable. The optimal tuning parameter used in our work was selected according to the Bayesian information criterion (BIC). Figures A.1 and A.2 below report the number of selected variables using cross-validation and BIC, respectively, for each estimated part of the regularization path (each $\lambda$).

Recall that the Adalasso can be determined by the following equation:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\beta}} ||\boldsymbol{Y} - \boldsymbol{X}\beta||_2^2 + \lambda \sum_{p=1}^{P} w_p |\beta_p|, \tag{A-1}$$

where $\boldsymbol{\beta}$ is the $\boldsymbol{N}\ \boldsymbol{x}\ \boldsymbol{1}$ vector of parameters, $\boldsymbol{\lambda}$ is the first-step Lasso regularization parameter, $\boldsymbol{w}$ is a weight vector defined as $\boldsymbol{w_p} = |\hat{\boldsymbol{\beta}}_{\boldsymbol{Lasso}}|^{-\gamma}$, with $\boldsymbol{b_p^*}$ as a first-step Lasso estimator for $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ is a non-negative weighting parameter ($\boldsymbol{\gamma > 0}$).

The parameters estimated using the Lasso were used as first-step estimates for the Adalasso's weighting vector ($\boldsymbol{w}$). We also set its weighting parameter $\boldsymbol{\gamma}$ at 1.

Figure A.1: LASSO's Tuning Parameter Selection via Cross-Validation

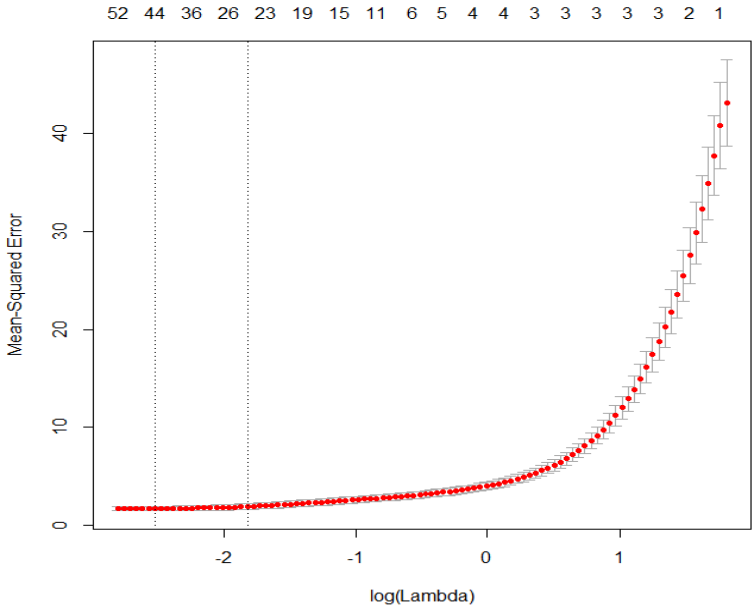
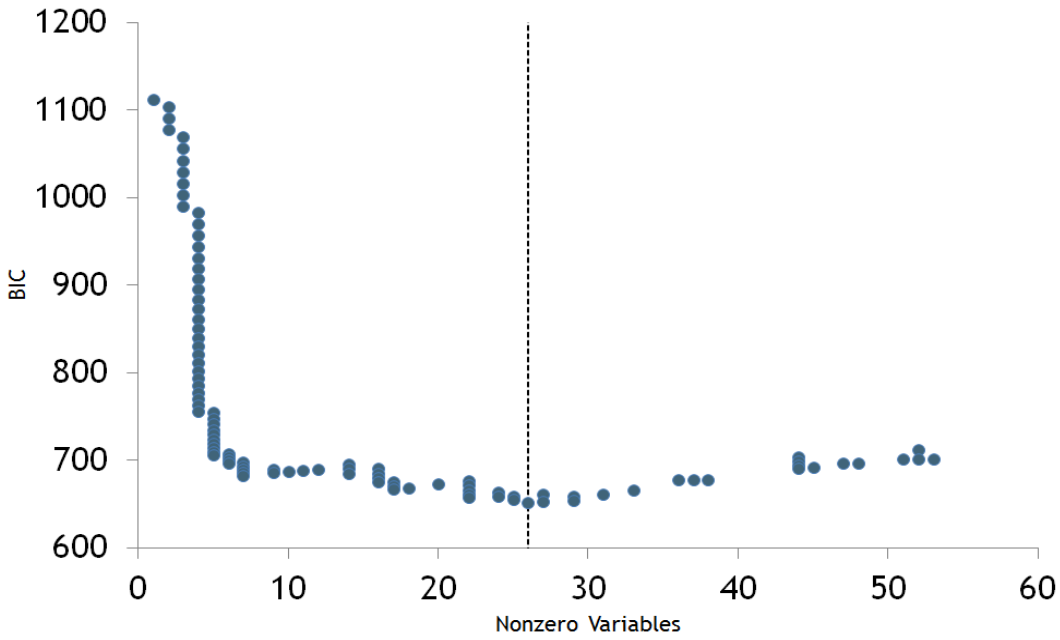





Figure A.2: LASSO's Tuning Parameter Selection via BIC



# B
# Data

From table B.1 to table B.14 we listed the name of each candidate variable, its label (the name used for each respective variable in our database), its source, and in which group it belongs (marked with an X). The sources listed below are further described in table B.15.

## B.1
## Candidate Variables

Table B.1: Industrial Sector

| Variable | Label | Source | Group 1 | Group 2 | Group 3 |
|---|---|---|---|---|---|
| Cardboard Expedition | ABPO | BBG | X | X | X |
| Heavy Vehicles Traffic | ABCR | BBG | X | X | X |
| Energy Consumption | ONS | BBG | X | X | X |
| Vehicle Stocks | Estoques_anfavea | ANFAVEA | X | X | X |
| Vehicle Production | Anf_totalsemrevisao | ANFAVEA | X | X | X |
| FIESP Industrial Index | INA_semrevisao | FIESP | X | X | X |

Table B.2: Retail

| Variable | Label | Source | Group 1 | Group 2 | Group 3 |
|---|---|---|---|---|---|
| Consumption Survey - Composite | SERASA | SERASA | - | X | X |
| Consumption Survey - Home appliances | SERASA_eletro | SERASA | - | X | X |
| Consumption Survey - Clothing | SERASA_vest | SERASA | - | X | X |
| ANFAVEA - Auto sale | ANFAVEA_vendas | ANFAVEA | - | X | X |
| Credit Card Sales | Cielo | Cielo | - | X | X |

Table B.3: Surveys: Current Conditions

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Manufacturing Industry | Prod_ISA | FGV | X | X | X |
| Consumption Goods | benscon_ISA | FGV | - | - | X |
| Durable Goods | bensd_ISA | FGV | - | - | X |
| Non-Durable Goods | bensnd_ISA | FGV | - | - | X |
| Intermediary Goods | bensint_ISA | FGV | - | - | X |
| Capital Goods | bensk_ISA | FGV | - | - | X |

Table B.4: Surveys: Expectations

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Manufacturing Industry | Prod_IE | FGV | X | X | X |
| Consumption Goods | benscon_IE | FGV | - | - | X |
| Durable Goods | bensd_IE | FGV | - | - | X |
| Non-Durable Goods | bensnd_IE | FGV | - | - | X |
| Intermediary Goods | bensint_IE | FGV | - | - | X |
| Capital Goods | bensk_IE | FGV | - | - | X |

Table B.5: Surveys: Composite

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Manufacturing Industry | Prod_ICI | FGV | X | X | X |
| Consumption Goods | benscon_ICI | FGV | - | - | X |
| Durable Goods | bensd_ICI | FGV | - | - | X |
| Non-Durable Goods | bensnd_ICI | FGV | - | - | X |
| Intermediary Goods | bensint_ICI | FGV | - | - | X |
| Capital Goods | bensk_ICI | FGV | - | - | X |

Table B.6: Other Surveys

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Manufacturing Industry - Total Demand | ind_FGV_trfo_DT | FGV | X | X | X |
| Manufacturing Industry - Capacity Utilization Index | ind_FGV_trfo_UCI | FGV | - | - | X |
| Manufacturing Industry - Expected Level of Production | ind_FGV_trfo_PP | FGV | - | - | X |
| Manufacturing Industry - Expected Level of Demand | ind_FGV_trfo_DTP | FGV | - | - | X |
| Manufacturing Industry - External Demand | ind_FGV_trfo_DE | FGV | - | - | X |



Table B.7: Foreign Trade

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Total Exports | Funcex_Xtotal | FUNCEX | X | X | X |
| Manufacturing Exports | Funcex_XManu | FUNCEX | X | X | X |
| Total Imports | Funcex_MTotal | FUNCEX | X | X | X |
| Industrial Imports | Funcex_II | FUNCEX | X | X | X |
| Imported Capital Goods parts | Funcex_peças | FUNCEX | X | X | X |
| Imported Vehicle equipment | Funcex_transp | FUNCEX | X | X | X |
| Imported Capital Goods | Funcex_bensk | FUNCEX | X | X | X |

Table B.8: Credit

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Concessions - Composite | BCB_conc | BCB | - | X | X |
| Concessions - Corporate | BCB_conPJ | BCB | - | X | X |
| Concessions - Non-Corporate | BCB_conPF | BCB | - | X | X |
| Interest Rate - Composite | BCB_juros | BCB | X | X | X |
| Interest Rate - Corporate | BCB_jurosPJ | BCB | X | X | X |
| Interest Rate - Non-Corporate | BCB_jurosPF | BCB | X | X | X |
| Spread Rate - Composite | BCB_spread | BCB | - | X | X |
| Spread Rate - Corporate | BCB_spreadPJ | BCB | - | X | X |
| Spread Rate - Non-Corporate | BCB_spreadPF | BCB | - | X | X |



Table B.9: Fiscal

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Total Tax Collections | fiscal_arrecadacao | NTD | X | X | X |
| Industrial Production Tax Collections | fiscal_IPI | NTD | X | X | X |
| Imports Tax Collections | fiscal_II | NTD | X | X | X |

Table B.10: Other Variables

| **Variable** | **Label** | **Source** | **Group 1** | **Group 2** | **Group 3** |
|---|---|---|---|---|---|
| Working Days | DU | MDIC | X | X | X |

## B.2 Sources

Table B.11: Candidate Variable Sources

| **Source Label** | **Description** |
|---|---|
| ABRAS | Brazilian Supermarket Association |
| BBG | Bloomberg |
| BCB | The Brazilian Central Bank |
| CIELO | Cielo Enterprise |
| FGV | Getúlio Vargas Foundation |
| FIESP | Industry Federation of the state of São Paulo |
| FUNCEX | Foreign Trade Studies Foundation |
| MDIC | The Ministry of Development, Industry and Foreign Trade and the Ministry of Labour |
| NTD | The National Treasury Department |
| SERASA | Serasa Experian Enterprise |

# C
# Figures

Figure C.1: The PIM and the Industrial GDP annual growth

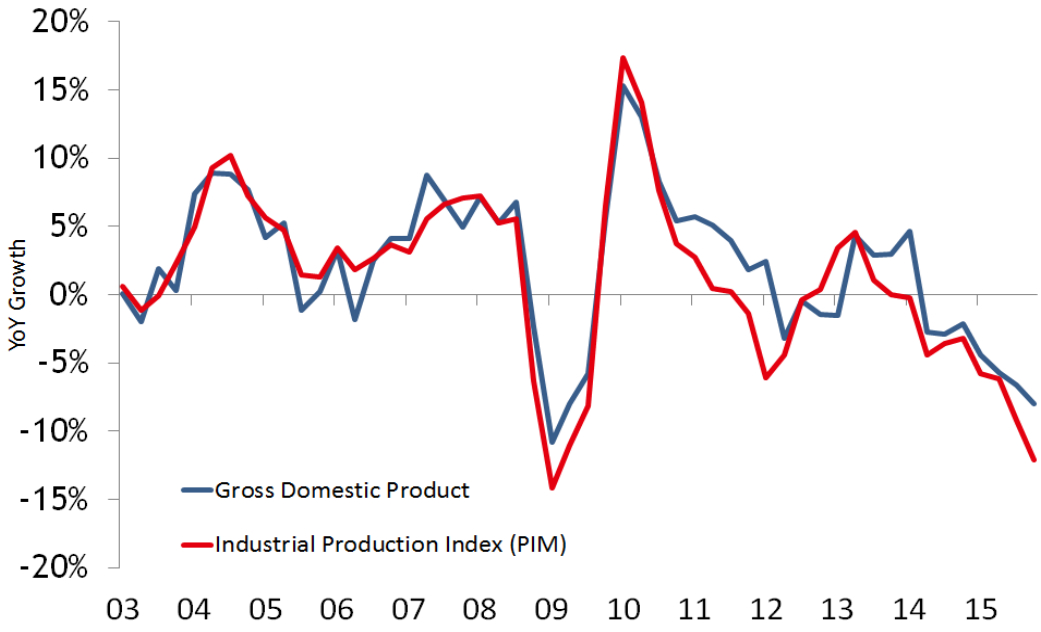


Figure C.2: LASSO Regularization Path

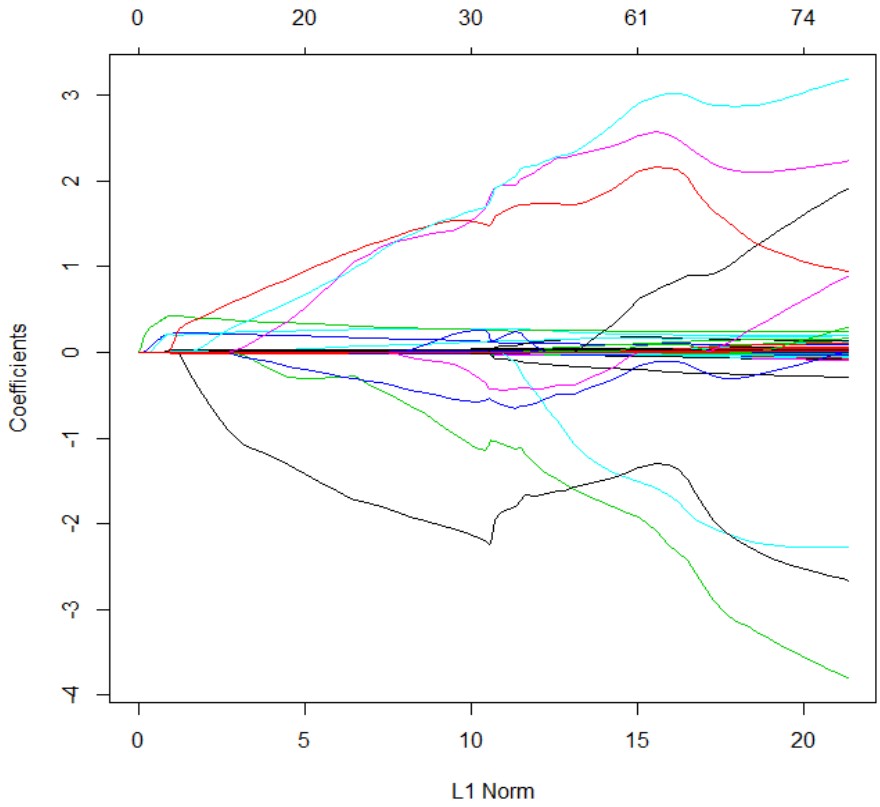



Figure C.3: The PIM's Average Seasonality

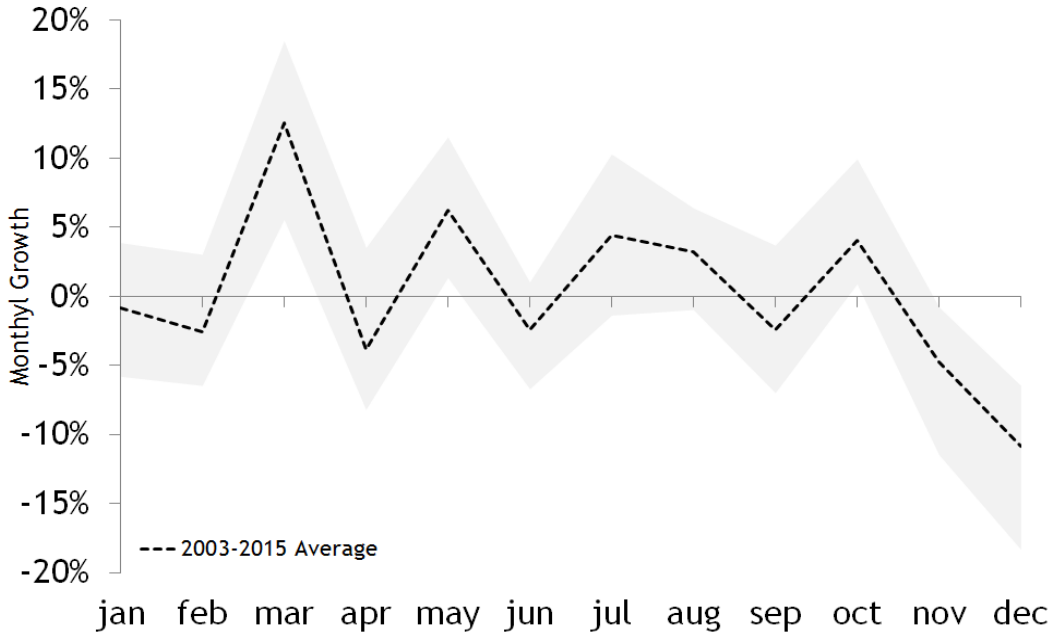


Figure C.4: The PIM's Seasonal Factors

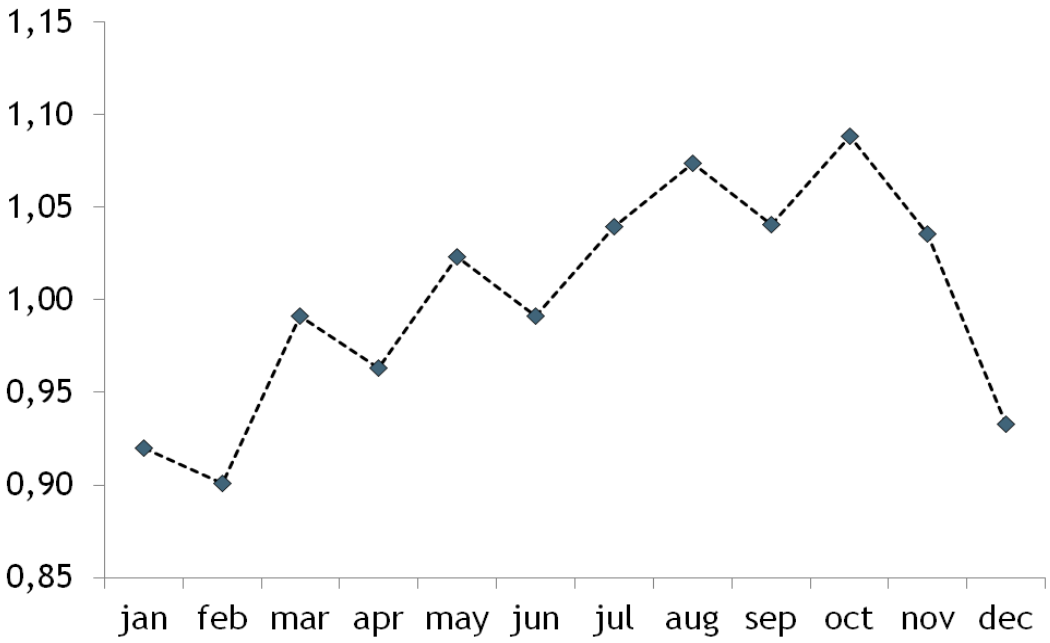



Figure C.5: One-step ahead Forecasting Performance

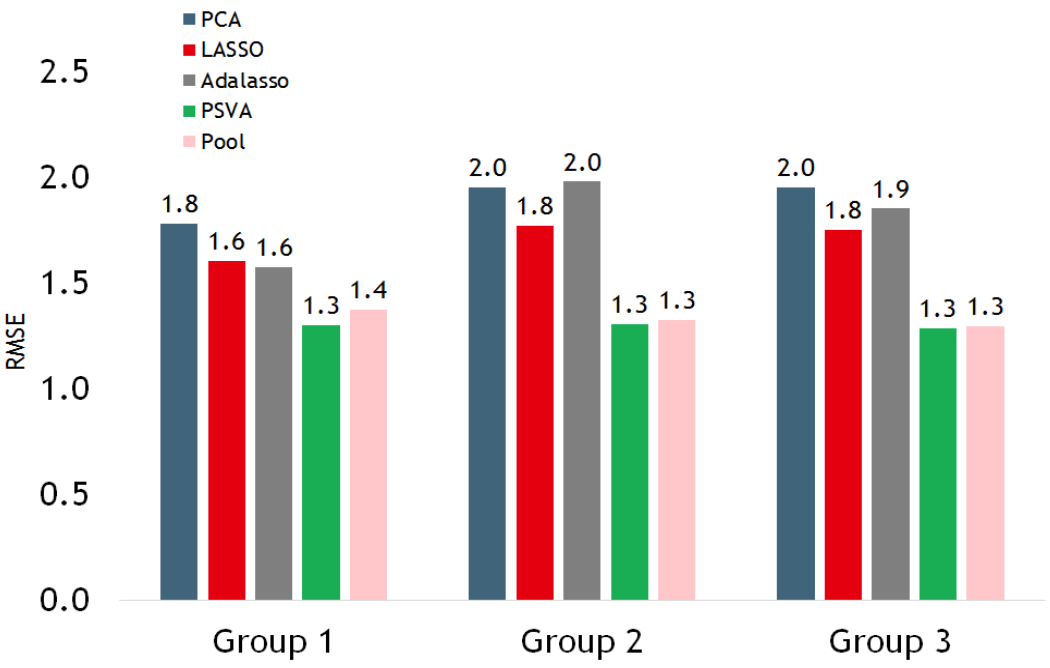


Figure C.6: Average Number of Selected Variables - Group 1

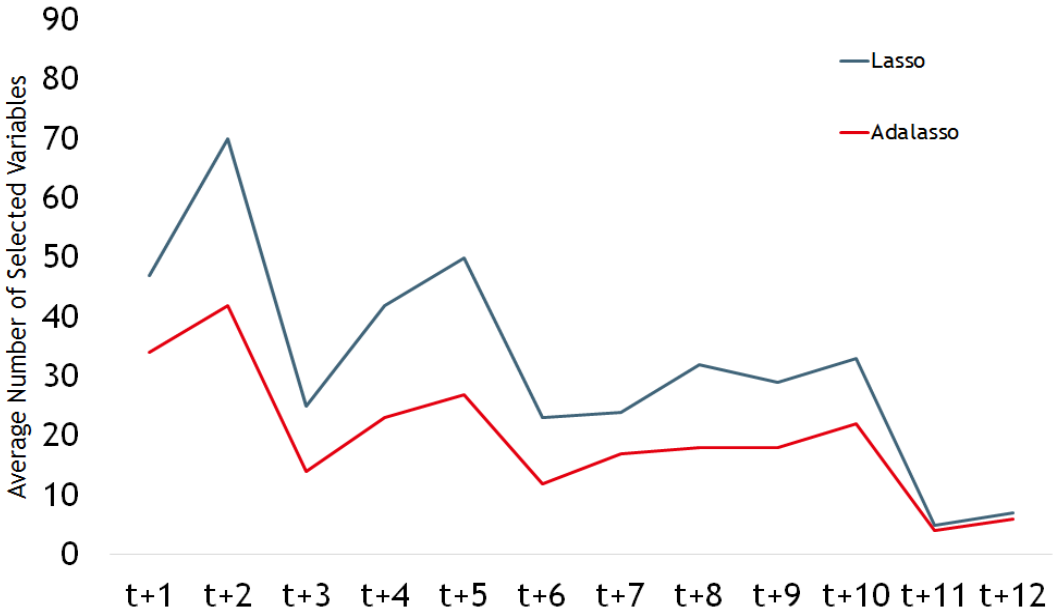



Figure C.7: Average Number of Selected Variables - Group 2

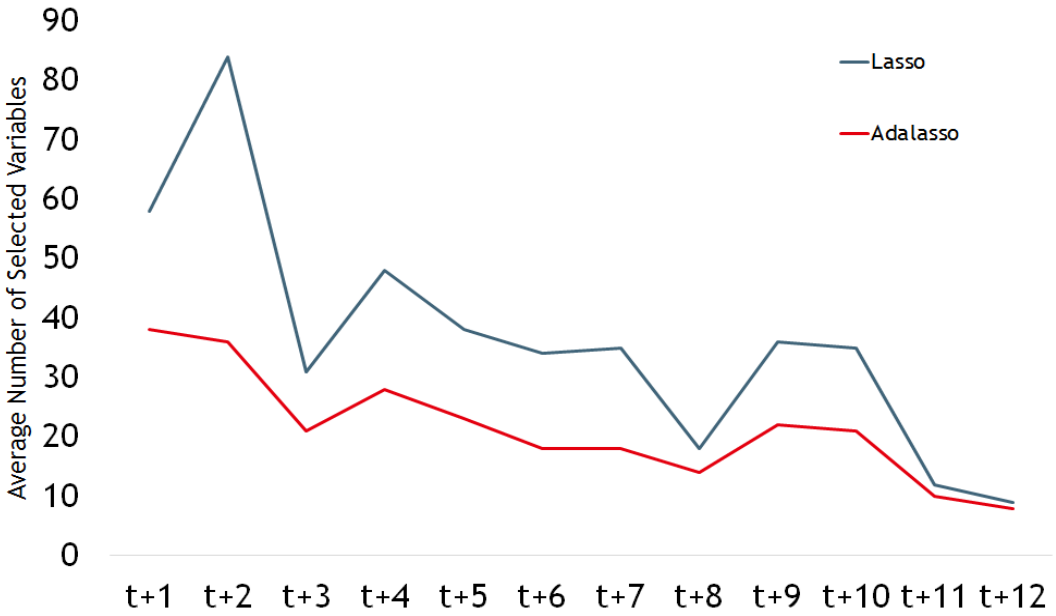


Figure C.8: Average Number of Selected Variables - Group 3

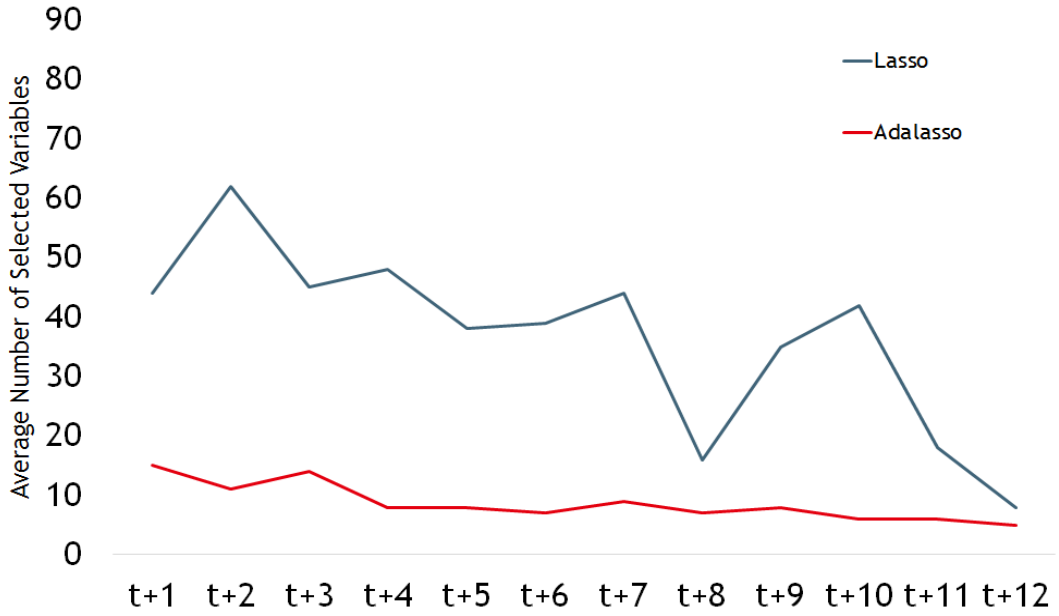



Figure C.9: Relative Performance to BBG Median Forecast RMSE

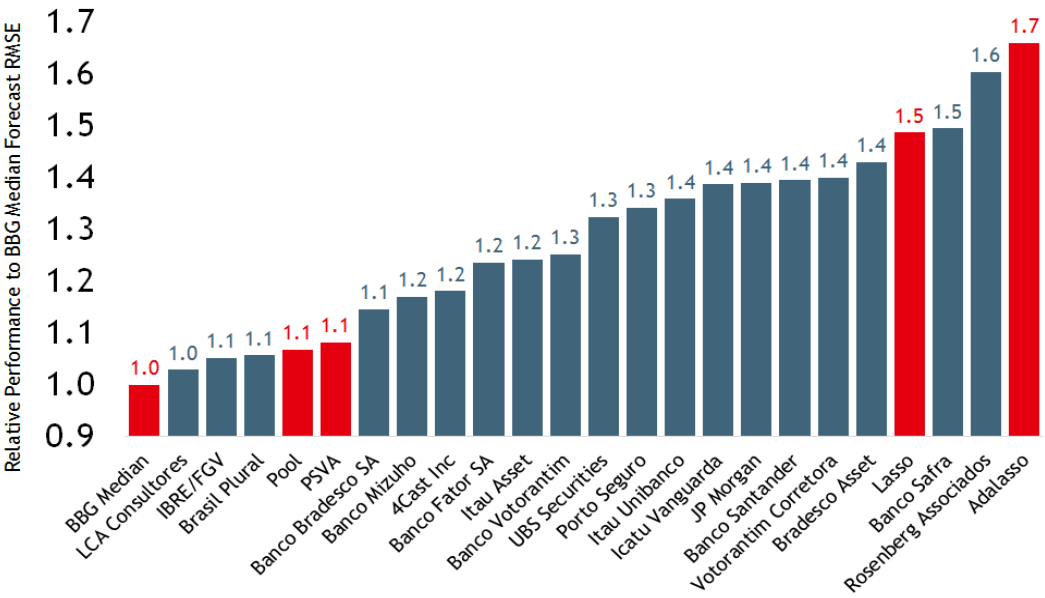


Figure C.10: Root Mean Squared Error

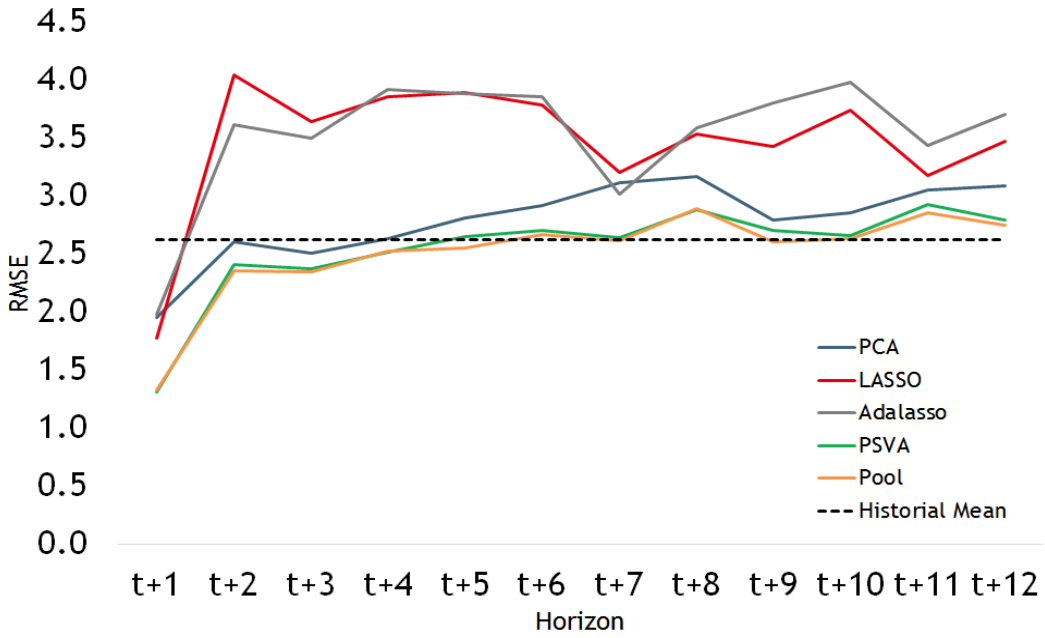



Figure C.11: Mean Absolute Error