Automated model selection with applications to Brazilian Industrial Production index *

Jordano Vieira Rocha[†]
Pedro L. Valls Pereira[‡]

October 23, 2019

dateSeptember, 2019

Abstract

Brazilian Industrial Production Index undergoes different methodological updates and periods of high inflation over time, which prompts researchers to avoid using too long industrial production series. We analyze how performance of different models in forecasting the Brazilian Industrial Production Index one-step ahead is influenced by the use of samples of different lengths. Relative performance of these models is also assessed. Results show that most models benefit from expanding the estimation sample beginning at least up to 1993:12. Autometrics lag selection with impulse dummy saturation forecasting performance is improved almost monotonically with sample size. For estimation starting in January 1975 and 1985, predictions from Autometrics with impulse dummy saturation and the average of forecasts are statistically more accurate than those from the benchmark AR model. However, the average of predictions performs better in the first half of the forecast horizon and Autometrics performs better in the second half.

Keywords: industrial production index, nonlinear methods, lag selection, dummy saturation, forecasting.
JEL Classification: C22, C52, C53.

1 Introduction

Macroeconomic data are an indispensable input to policy makers and private agents in capturing the economic situation of an entire country or region. They are intended to translate a complex set of heterogeneous variables into informative simple statistics. Since collecting information on all economic agents is unfeasible, macroeconomic data are usually originated from periodic surveys and a limited number of administrative records.

*The first author acknowledges financial suppport form CAPES and IBGE and second author from CNPq (309158/2016-8) and FAPESP (2013/22930-0).
† IBGE
‡ Sao Paulo School of Economics - FGV and CEQEF -FGV

Given technological advances and changes in the structure of the economy, the methodology of such surveys are updated over time. Therefore, using long macroeconomic series may be problematic, specially in forecasting, since recent and old data may be of different qualities. Additionally, old data may contain information that is no longer relevant in forecasting future values (this is a characteristic not only of macroeconomic data, but of socioeconomic time series in general).

To investigate the effects of sample size for forecast performance, we analyze the Brazilian industrial production series. The main index of industrial production in Brazil is the Monthly Industrial Survey - Physical Production (PIM-PF), produced by the Brazilian Institute of Geography and Statistics (IBGE). This is one of the longest monthly macroeconomic time series for Brazil, with initial observations in 1975. It is widely used as a business cycle indicator, since the GDP is disclosed quarterly, while the PIM-PF is a monthly index.

Although, in its essence, the index is practically the same over the years — a weighted average of relative quantities, developed over a Laspeyres *quantum* index — this series has undergone methodological benchmark revisions over time, which consisted, basically, in updating the weights for products and activities, the surveyed set of products and respondents, the level of regional disaggregation, and the industry activity and product classifications, following revisions in international standards.

The last benchmark revision was made in March 2014, which produced revised values up to January 2002. This current series may be insufficiently large, particularly when a fraction of the sample is used for forecast evaluation. On the other hand, as stated before, larger series may contain irrelevant information, suffer from the different methodologies and from the presence of outliers, which may harm parameter estimation. Furthermore, since weights for activities composing the index are fixed for each methodological benchmark, periods of high inflation tend to have distorting influence over the series, possibly generating additional noise.

One objective of this paper is to study the effect of the sample size selection in the forecast performance. Given the intrinsic volatile behavior of industrial production and the structural breaks that may arise also due to changes in the survey methodology, a second objective is to evaluate different linear and nonlinear models, comparing relative performance between them using different sample sizes. We consider only univariate models to rule out the influence of other variables in the analysis.

It is a standard assumption that economic time series are realizations of stochastic processes, represented by a Data Generating Process (DGP)[1]. The DGP (or at least a close approximation) is assumed to be known and any abnormal shift, departing from the standard linear dynamics, in the series must be captured by its form.

Nonlinear models use different techniques to capture these kind of break in a series over time. Some of the most known are composed by a set of different linear models that are alternated in time according to some rule of transition. It is the case of the Markov Switching Autoregressive Model (MSAR), which have a transition rule based on a probabilistic struc-

[1]Some authors consider that the DGP represents the joint density, not necessarily with constant parameters, of all variables characterizing an economy. Following the Theory of Reduction (Hendry, 1995), a subset of these variables can be represented by a Local Data Generating Process (LDGP). To properly parametrize any time-varying effect, an econometric model, usually different from the LDGP, must be a function with constant parameters. See, for example, Bontemps and Mizon (2003).

ture, and the Smooth Transition Autoregressive Model (STAR), with a changing rule based on a transition function that depends on the value of the variable of interest. STAR transitions tend to be more abrupt than those of MSAR, because, in the latter, the transitions are ruled by a Markov Process. This process is characterized by a transition matrix which give the probabilities of switching between regimes, and this characterizes the smoothness of transitions. Early examples of MSAR and STAR models application can be found in Engel (1994) and Teräsvirta and Anderson (1992), analyzing the the US dollar exchange rate and the quarter industrial production indexes from several countries, respectively.

Another way to capture nonlinearities, without explicitly modelling their behavior, is to automatically detect outliers and level shifts, using impulse and step dummies in the model. Santos et al. (2008) propose this approach, selecting the relevant dummies based on the Autometrics algorithm of variable selection (Doornik, 2009). Correctly modeled impulse and step dummies can improve parameter estimation and detect a change in the series' mean, modifying the conditional mean and avoiding repeatedly wrong forecasts, which would tend to be biased towards the previous unconditional mean. Hendry and Mizon (2011), for example, study the food expenditure in United States using this technique.

However, econometric models may still fail in the presence of unpredictable structural breaks (or changes in the underlying stochastic process). Such fact have been gaining more attention specially after the subprime financial crisis. Abrupt changes in the dynamic of several economic and financial series have caused a phenomenon called forecast failure (*e.g.*, Hendry, 2012), *i.e.*, persistent failures in the forecasts after a certain period.

Defining robust forecasting mechanisms, even in the presence of structural breaks of unknown form, becomes important in a world with abrupt, unpredictable events. Hendry (2006) suggests that to predict using the difference of a series (even a stationary one) would make its forecast robust. In the moment of a structural break occurrence, the forecast would be biased, since this change would be unpredictable by definition. In the following periods, however, the forecast would become unbiased, although at the cost of a higher forecast error variance in times of economic stability.

Besides the aforementioned methods, we analyze forecasts from models selecting lags of the dependent variable: the benchmark Autoregressive Model of order p (chosen by the Bayesian Information Criterion), against which all other forecasts will be compared, the Autometrics algorithm without any dummy saturation, and the Least Absolute Shrinkage and Selection Operator (LASSO) and two of its variants, the Adaptive LASSO (AdaLASSO) and the Weighted Lag Adaptive LASSO (WLAdaLASSO), well known regularization and variable selection methods (Tibshirani, 1996; Zou, 2006; Konzen and Ziegelmann, 2016). The simple average of forecasts from all models are also computed.

We perform Diebold and Mariano (1995) test to assess the relative forecast performance using different sample sizes in relation to the current smaller sample (starting in January 2002) for each method. We also test, for each sample, if these methods' forecasts are more accurate than the benchmark AR(p) model within each sample. Forecasts are made one-step ahead, re-estimating and re-specifying the models at each step. Results show that most models benefit from enlarging the estimation sample at least to 1993:12 and that a rolling window forecasting scheme with this same initial sample performs worse than its expanding window analogous, pointing that information in that period may remain relevant in forecasting future values.

In addition to this introduction, this paper is structured as follows. Section 2 contains an overview of the PIM-PF and the nature of its methodological revisions, Section 3 provides a theoretical review of the considered models, Section 4 details the data used and the settings of the forecasting exercise performed, Section 5 presents the results, Section 6 contains the concluding remarks.

2 An Overview of the Monthly Industrial Survey and its Methodological Revisions

The Monthly Industrial Survey is part of an integrated system of business surveys[2], consisting of annual (or structural) and monthly (or conjunctural) surveys — the annual and monthly surveys of Mining and Manufacturing (PIA and PIM-PF), of Trade (PAC and PMC), and of Services (PAS and PMS) and the annual survey of Construction (PAIC). There are, additionally, satellite (or special) surveys, with specific themes, for example, the Innovation Survey (formerly Industrial Survey of Technological Innovation, or PINTEC).

This system is constructed over a unified business register, named Central Business Register (CEMPRE), which was built in the major revision of business surveys carried out in the mid 1990s. The CEMPRE consists of data and industry classifications of all formal enterprises and their local units, *i.e.*, all units registered in the Internal Revenue Service, identified by a unique National Register of Juridical Person (CNPJ) number. It is used to construct the samples of the annual surveys, which, in turn, serve as a basis to calculate the weights and extract the samples of monthly surveys.

Regarding the PIM-PF, the selection of products and respondents of the survey is based on a panel of units representing a percentage of the Industrial Transformation Value (VTI) or the Gross Value of Industrial Production (VBPI) on a base year. In the current benchmark, for example, in each region, only those activities representing 80% of the VTI are surveyed. Of those activities, only the products representing 80% of the VBPI are selected. The local units responsible for 70% of production for each product are surveyed. These fractions are relative to the base year of 2010.

The methodological revisions must be performed periodically to update the set of products and respondents surveyed, the activities analyzed, the industrial classifications, the weighting structure, based on more timely information from economic censuses or annual surveys, and the disagreggation of the index in terms of geographical regions. Seasonal adjustment procedures are also updated as new methodological techniques are developed and suggested by international standards.

The index *per se* is basically the same, simply a weighted average of relative quantities, developed over a Laspeyres *quantum* index:

[2] This section is based on IBGE's methodological reports (IBGE, 1991, 1996, 2004, 2015b), on information from PIM-PF monthly publications and IBGE's metadata website — available, respectively, at https://metadados.ibge.gov.br and at https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=7228 — and on Góes (2005).

$$I_t^0 = \sum_{i=1}^{n} w_i^0 \frac{q_i^t}{q_i^0}, \tag{1}$$

where I_t^0 is the index in time t relative to the base year 0, $w_i^0 = (VTI_i^0 / \sum_{i=1}^{n} VTI_i^0)$, and q_i^t/q_i^0 is the quantity[3] of product i in time t relative to the base period 0 (*e.g.*, 2012 in the current series)[4].

The first statistics of industrial physical production in Brazil were made in the early 1970s. Initially, the survey included 110 products and around 1000 respondents, using weights from the then recently developed Annual Industrial Survey of 1968. The first reformulation was made with the 1970 economic census publication, available in 1975, which included 600 products and 2500 respondents.

The following update in the series used weights from the 1978 PIA and information from the 1980 Industrial Census, including 736 products and 5000 respondents. With the 1985 Industrial Census (the last economic census to be carried out), the new PIM-PF structure comprised 944 products and 6200 respondents.

The economic statistics system described in the beginning of this section started to be implemented in the mid 1990s. Before this framework, the Industrial Census worked both as a register and provided information for the weighting structure and selection of product and respondent samples, with additional information from the Annual Surveys. The register based on the censuses was rapidly out of date, due to the lag which the census results were available, aggravated by the high rates of birth and deaths of enterprises. The CEMPRE, which is updated each year with administrative data and results from the most recent surveys, was developed to overcome those difficulties. Other important technical changes were the use of a new industrial classification system of activities, the National Classification of Economic Activities (CNAE — IBGE, 2003), and of products (PRODLIST-Indústria), and the redefinition of the statistical unit of investigation, from establishment to local unit[5].

These changes were incorporated in the PIM-PF only in 2004, covering data from January 2002, which used information from the 1998-2000 editions of the PIA, including 830 products and 3700 local units. The last methodological revision, made in 2015, covering data starting from January 2012, implemented the changes to make the survey compatible with second version of CNAE (IBGE, 2015a). It selected 944 products and 7800 local units.

[3]The relative quantities are actually calculated as $\frac{q_i^t}{q_i^0} = \left(\frac{q_i^1}{q_i^0}\right)^* \left(\frac{q_i^2}{q_i^1}\right)^* \cdots \left(\frac{q_i^t}{q_i^{t-1}}\right)^*$, where $\left(\frac{q_i^t}{q_i^{t-1}}\right)^*$ is the relative quantity between months t and $t-1$ of product i, maintaining the same intersection of respondent and product panels from months $t-1$ and t. Note that $\left(\frac{q_i^t}{q_i^{t-1}}\right)^* \neq \frac{q_i^t}{q_i^{t-1}}$, since panels may be different in t and $t-1$. This adjustment is done to account for this fact, ensuring that relatives are calculated in compatible panels.

[4]In some editions, the weights are calculated over the Value Added instead of the Industrial Transformation Value, *i.e.*, $w_i^0 = (VA_i^0 / \sum_{i=1}^{n} VA_i^0)$, such as in the series that use weights from the 1985 Industrial Census.

[5]The establishment is defined as an enterprise or part of enterprise that, in a given location, independently engages predominantly in one economic activity. Thus one enterprise may have multiple establishments. The local unit is defined as a physical space where one or more economic activities are developed, corresponding to a unique CNPJ (14 digits) number. The use of establishment as statistical unit revealed problematic, since respondent had difficulties filling out questionnaires and it brought additional complexity in the collecting and processing of information. Furthermore, the harmonization between statistical and legal definition, together

with a common industrial classification, made it easier to use administrative data in the register and surveys.

2.1 The Chaining of Series from Different Methodological Benchmarks

Each methodological update also changes the base year, *i.e.*, the year in which the average of the index is 100. This results in multiple series with different levels. For the sake of the series continuity, if we want to obtain a longer time length containing all available data, the entire period must be expressed in the same base.

IBGE's last technical report (IBGE, 2015b) describe how to link two consecutive series with different base periods. Although this is done to the series of all economic activities that compose the index, we made this procedure only to the General Index, since we do not work with disaggregated data.

Suppose we want to build a series $\{I_t\}_{t=1}^{K}$ of an index. Let $b_1 = (k_{b_1}^{Jan}, k_{b_1}^{Feb}, \dots, k_{b_1}^{Dec})$ and $b_2 = (k_{b_2}^{Jan}, k_{b_2}^{Feb}, \dots, k_{b_2}^{Dec})$ be the sets of months composing two different base years. Suppose that we have two different monthly series of the index expressed in the base years b_1 and b_2, $\{I_t^{b_1}\}_{t=1}^{k_1}$ and $\{I_t^{b_2}\}_{t=k_2}^{K}$, $k_2 < k_1$, and that $b_2 \subseteq (k_2, k_2+1, \dots, k_1)$. This states that, to make the link between the series possible, both series must have observations in a given base year. The definition of base year implies that

$$\frac{\sum_{t \in b_1} I_t^{b_1}}{12} \equiv 100 \text{ and } \frac{\sum_{t \in b_2} I_t^{b_2}}{12} \equiv 100. \tag{2}$$

The index of base b_1 expressed in base b_2 is given by

$$I_t^{b_1|b_2} = I_t^{b_1} \cdot \frac{\sum_{t \in b_2} I_t^{b_2}}{\sum_{t \in b_2} I_t^{b_1}}. \tag{3}$$

Note that

$$\frac{\sum_{t \in b_2} I_t^{b_1|b_2}}{12} = \frac{\sum_{t \in b_2} I_t^{b_1}}{12} \cdot \frac{\sum_{t \in b_2} I_t^{b_2}}{\sum_{t \in b_2} I_t^{b_1}} = 100. \tag{4}$$

The entire series can be, thus, expressed in terms of b_2:

$$\{I_t\}_{t=1}^{K} \equiv \left(I_1^{b_1|b_2}, I_2^{b_1|b_2}, \dots, I_{k_2-1}^{b_1|b_2}, I_{k_2}^{b_2}, I_{k_2+1}^{b_2}, \dots, I_K^{b_2} \right). \tag{5}$$

3 Theoretical Review

This section briefly describes the models used in this paper to forecast the Brazilian Industrial Production Index series. In following sections, forecast performance of these techniques will be compared to the benchmark Autoregressive Model of order p. We consider non-linear methods (the Markov Switching and Smooth Transition Autoregressive Models), an automatic lag selection and outlier detection method (Autometrics), a naive method robust

to structural breaks (Double Difference Device), and Lasso-type penalty methods (LASSO, Adaptive LASSO and Weighted Lag Adaptive LASSO).

3.1 Markov Switching Autoregressive Model

This technique supposes that the DGP consists in different autoregressive models which alternate between them according to a process represented by a discrete latent variable $s_t = \{1, 2, \ldots, N\}$, in which each value represents a different regime. An additional assumption is that s_t follows a Markov process, *i.e.*, the probability of the realization of state j in the current period depends only of the realization of the state from the immediately previous period:

$$P(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots) = P(s_t = j | s_{t-1} = i) = p_{ij}. \tag{6}$$

The set of probabilities of transition p_{ij} can be represented, thus, by a transition matrix P:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix} \tag{7}$$

Suppose, for simplicity and without loss of generality, an AR(1) process in which the intercept (c_{s_t}) and autoregressive values (φ_{s_t}) change according to each regime. We can write this process as:

$$y_t = c_{s_t} + \varphi_{s_t} y_{t-1} + \varepsilon_t, \tag{8}$$

where we assume that $\varepsilon_t \sim NI(0, \sigma^2_{s_t})$, *i.e.*, the variance of the error term also depends on s_t.

Observe that, even if we know all the involved parameters, we cannot be sure of which regime the process is in each period. Let θ be the vector containing all the parameters in the model. We can infer the state by the probability that the process is in regime j conditional to the realization of the series values up to t, denoted $P(s_t = j | y_t, \theta)$. Following the law of conditional probabilities:

$$P(s_t = j | y_t, \theta) = \frac{p(y_t, s_t = j, \theta)}{f(y_t, \theta)} = \frac{\pi_j f(y_t | s_t = j, \theta)}{f(y_t, \theta)}. \tag{9}$$

In the expression above, $\pi_j = P(s_t = j, \theta)$ is the unconditional probability of s_t assuming the value j, $f(y_t | s_t = j, \theta)$ is the conditional density, assumed normal, that is:

$$f(y_t | s_t = j, \theta) = \frac{1}{\sqrt{2\pi}\sigma_{s_t}} exp \left\{ \frac{-(y_t - c_{s_t} - \varphi_{s_t} y_{t-1})^2}{2\sigma^2_{s_t}} \right\} \tag{10}$$

and the unconditional density is given by:

$$f(y_t, \theta) = \sum_{j=1}^{N} p(y_t, s_t = j, \theta) = \sum_{j=1}^{N} \frac{\pi_j}{\sqrt{2\pi}\sigma_{s_t}} exp \left\{ \frac{-(y_t - c_{s_t} - \varphi_{s_t} y_{t-1})^2}{2\sigma^2_{s_t}} \right\}. \tag{11}$$

Stacking the conditional densities $f(y_t | s_t = j, \theta)$, the conditional probabilities $P(s_t = j | y_t, \theta)$ and the forecasts $P(s_{t+1} = j | y_t, \theta)$ in vectors $(N \times 1)$, respectively, η_t, $\hat{\xi}_{t|t}$ e $\hat{\xi}_{t+1|t}$, *i.e.*,

$$\eta_t = \begin{bmatrix} f(y_t|s_t = 1, y_{t-1};\theta) \\ f(y_t|s_t = 2, y_{t-1};\theta) \\ \vdots \\ f(y_t|s_t = N, y_{t-1};\theta) \end{bmatrix} \tag{12}$$

$$\hat{\xi}_{t|t} = \begin{bmatrix} P(s_t = 1|y_t, \theta) \\ P(s_t = 2|y_t, \theta) \\ \vdots \\ P(s_t = N|y_t, \theta) \end{bmatrix} \tag{13}$$

$$\hat{\xi}_{t+1|t} = \begin{bmatrix} P(s_{t+1} = 1|y_t, \theta) \\ P(s_{t+1} = 2|y_t, \theta) \\ \vdots \\ P(s_{t+1} = N|y_t, \theta) \end{bmatrix}. \tag{14}$$

Hamilton (1994) shows that the optimal inference for the regimes is given by the following recursion:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{1'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{15}$$

$$\hat{\xi}_{t+1|t} = P\hat{\xi}_{t|t}, \tag{16}$$

where P is defined by (7), $\odot$ represents element-by-element multiplication and 1 is a $(N \times 1)$ vector of 1's.

The likelihood maximization problem is given, thus, by:

$$max\ L(\theta) = \sum_{t=1}^{T} f(y_t|y_{t-1}, \theta) = \sum_{t=1}^{T} 1'(\hat{\xi}_{t|t-1} \odot \eta_t) \tag{17}$$

$$s.t.\ (p_{i1} + p_{i2} + \cdots + p_{iN}) = 1, p_{ij} \geq 0, \forall j, i. \tag{18}$$

Expression (15) is similar to (9), and (17) similar to (11).

Given the values of θ_0 and $\hat{\xi}_{1|0}$ (*e.g.* $\hat{\xi}_{1|0} = \rho$, with $\rho = 1/N$), the recursion in (15)-(16) returns the values of $\hat{\xi}_{t|t}$. Note that, for each t, this method utilizes only the realized values in the current and in previous periods $(t, t-1, t-2, \dots)$ in order to estimate $P(s_t = j|y_{t-1}, \theta)$. However, we possess additional information to increase this estimate's accuracy, namely, the remainder of the sample $(t+1, t+2, \dots, T)$. For this, we have to use a smoothed inference mechanism.

Let us denote, generalizing the previous notation, the probability of being in a regime conditional to the realized values from the series and to the parameters $P\,(s_t = j \mid y_\tau\,,\,\theta)$. For $t < \tau$, we will have a smoothed inference of regime t. Kim (1994) develops a recursion analogous to (16):

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \left\{ P' \left[\hat{\xi}_{t+1|T} (\div) \hat{\xi}_{t+1|t} \right] \right\}, \tag{19}$$

where $(\div)$ means term-by-term division.

The iteration algorithm starts in $\hat{\xi}_{T|T}$, which, in turn, is calculated from (15)-(16), passing through $\hat{\xi}_{T-1|T}, \hat{\xi}_{T-2|T}, \ldots, \hat{\xi}_{1|T}$. Once having $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$ known, the full vector of maximum likelihood estimates $\hat{\theta}$ is obtained solving (17)-(18), proceeding recursively until there is convergence to a θ^*. Restricting $p_{ij} \geq 0$ and $p_{i1} + p_{i2} + \cdots + p_{iN} = 1$ for all i, j, and supposing $\hat{\xi}_{1|0}$ is fixed and not related to the other parameters, Hamilton (1990) demonstrate the the transition probabilities estimators p_{ij} are given by:

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} P(s_t = j, s_{t-1} = i \,|\, y_T, \hat{\theta})}{\sum_{t=2}^{T} P(s_{t-1} = i \,|\, y_T, \hat{\theta})}, \tag{20}$$

that is the sum of the probabilities of regime i being followed by regime j divided by the sum of the probabilities of being in regime i.

Finally, to calculate the optimal forecast one-step ahead, suppose we know which will be the regime in period $t+1$, so the prediction will be obtained by $h_{jt} = \mathrm{E}(y_{t+1} | s_{t+1} = j, \mathcal{Y}_t; \theta) = \int y_{t+1} f(y_{t+1} | s_{t+1} = j, \mathcal{Y}_t; \theta) dy_{t+1}$, where $\mathcal{Y}_t$ is a vector containing all y observations through date t. The unconditional expectation will be, thus, given by:

$$\begin{aligned}
\mathrm{E}(y_{t+1} | x_{t+1}, \mathcal{Y}_t; \theta) &= \int y_{t+1} f(y_{t+1} | x_{t+1}, \mathcal{Y}_t; \theta) dy_{t+1} \\
&= \int y_{t+1} \left\{ \sum_{j=1}^{N} p(y_{t+1}, s_{t+1} = j | x_{t+1}, \mathcal{Y}_t; \theta) \right\} dy_{t+1} \\
&= \int y_{t+1} \left\{ \sum_{j=1}^{N} [f(y_{t+1} | s_{t+1} = j, x_{t+1}, \mathcal{Y}_t; \theta) P(s_{t+1} = j | x_{t+1}, \mathcal{Y}_t; \theta)] \right\} dy_{t+1} \\
&= \sum_{j=1}^{N} P(s_{t+1} = j | x_{t+1}, \mathcal{Y}_t; \theta) \int y_{t+1} f(y_{t+1} | s_{t+1} = j, x_{t+1}, \mathcal{Y}_t; \theta) dy_{t+1} \\
&= \sum_{j=1}^{N} P(s_{t+1} = j \,|\, \mathcal{Y}_t; \theta) \mathrm{E}(y_{t+1} \,|\, s_{t+1} = j, x_{t+1}, \mathcal{Y}_t; \theta).
\end{aligned} \tag{21}$$

If we stack h_{jt} in a vector $(N \times 1)$ h_t, then:

$$\mathrm{E}(y_{t+1} | \mathcal{Y}_t; \theta) = h_t' \hat{\xi}_{t+1|t}. \tag{22}$$

$t+1$

Although the Markov Chain admits a linear representation, the optimal forecast one-step ahead is a nonlinear function of the of the observed variables, for $\xi_{t|t}$ depends on y_t in a nonlinear way. As pointed by Hamilton (1994), this implies that, if, inside a regime, an outlier (an observation little likely to be generated in this regime) is observed, the framework analyzed here has high probability of inferring that a change of state has occurred.

3.2 Smooth Transition Autoregressive Model

As it is the case with MSAR models, the STAR models consist in two or more regimes following different linear autoregressive processes. The transition dynamics from one regime to another is, though, different. The change between states is given by a transition function as described below.

Suppose a stochastic process represented by the following DGP:

$$y_t = \pi_{10} + \pi_1' z_t + (\pi_{20} + \pi_2' z_t) F(y_{t-d}) + \epsilon_t. \quad (23)$$

$F(y_{t-d})$ is called the transition function, it depends on the model's dependent variable, y_t, lagged by the parameter d, that will be estimated. The vector $z_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ contains the lagged series, (π_1, π_2) are parameter vectors with correspondent dimensions and (π_{10}, π_{20}) are scalars. The error term is assumed normal, independent and identically distributed, *i.e.*, $\epsilon_t \sim NI(0, \sigma^2)$. It is important to stress that ϵ_t is symmetric, so that rejections of the null hypothesis from linearity tests would come from the model parametrization, and not from the error term.

The F function is limited between 0 and 1. In case it is an indicator function $F(y_{t-d}) = I(y_{t-d} > c)$[6], the model is denoted *Self Exciting Threshold Autoregressive* (SETAR), in which, when y_{t-d} is greater than the threshold c, the regime is changed instantaneously. To establish a smooth transition between the regimes, we can use some continuous function between 0 and 1, so that the parameters of (23) vary, as function of y_{t-d}, between π_{10} and $(\pi_{10} + \pi_{20})$ for the intercept, and between π_1 and $(\pi_1 + \pi_2)$, for the autoregressive parameters. These values correspond to the two extreme regimes ($F = 0$ and $F = 1$, respectively).

There are two functions used predominantly in the literature. The first is the logistic function:

$$F(y_{t-d}) = (1 + exp[-\gamma(y_{t-d} - c)])^{-1}. \quad (24)$$

When the model uses the function above, it is called *Logistic Smooth Transition Autoregressive* (LSTAR). The inclination parameter γ measures the speed of transition from one regime to another, and the location parameter c is the threshold. Besides, when $\gamma \to \infty$, the model approximates a SETAR model.

LSTAR model has been used to model macroeconomic series with asymmetric behavior. Öcal and Osborn (2000) and Teräsvirta and Anderson (1992) use it to model industrial production series and Skalin and Teräsvirta (2002) for unemployment series.

The second function is the exponential function:

$$F(y_{t-d}) = 1 - exp(-\gamma(y_{t-d} - c)^2). \quad (25)$$

Using this function, we denote the model Exponential Smooth Transition Autoregressive (ESTAR). The interpretation given to the parameters γ and c is the same as in LSTAR. When $\gamma \to \infty$, however, F tends to $I(y_{t-d} = c)$.

[6]We will use, in this work, only two regimes, corresponding in theory to periods of expansion and periods of recession, but we could generalize for more regimes. In this case, the model would have, in the indicator function example, the form $y_t = \sum_{j=1}^{r} \alpha_j' z_t I(c_{j-1} < y_{t-d} < c_j) + \epsilon_t$, with c_j representing each *threshold*.

The ESTAR model is normally used in series with symmetric behavior. In this case, we would have one regime when y_{t-d} gets near the threshold and another when it gets far. An example of application is Arango and Gonzalez (2001), who models the Colombian inflation acceleration.

To verify if there are gains in formulating a nonlinear specification, we should do a linearity test. Besides, we will need this test to estimate the parameter d. There is, however, a caveat in performing this test, for there are many ways of defining linearity in (23)-(24)/(25).

When we test $H_{01} : \gamma = 0$ in (23)-(24), the model will be identified only under the alternative hypothesis, because π_{20} and π_2 will be able to assume any value. If, on the other side, we test $H_{01} : \pi_{20}, \pi_2 = 0$, γ and c will be able to assume any value. A similar reasoning applies to (23)-(25).

To contour this problem, Teräsvirta (1994), based on Davies (1977) and on Luukkonen et al. (1988), elaborates a test from a Taylor expansion around $\gamma = 0$, leading to the following auxiliary regression:

$$\hat{u}_t = \hat{w}'_{1t}\tilde{\beta}_1 + \sum_{j=1}^{p} \beta_{2j} y_{t-j} y_{t-d} + \sum_{j=1}^{p} \beta_{3j} y_{t-j} y^2_{t-d} + \sum_{j=1}^{p} \beta_{4j} y_{t-j} y^3_{t-d} + v'_t, \qquad (26)$$

where $\hat{u}_t$ are the residuals from the (linear) AR(p) model estimation, $\hat{w}'_{1t} = -(1, z'_t)'$ and $\tilde{\beta}_1 = (\beta_{10}, \beta'_1)'$. The null hypothesis of linearity is:

$$H'_0 : \beta_{2j} = \beta_{3j} = \beta_{4j} = 0, \; j = 1, \ldots, p. \qquad (27)$$

Each β_{ij} is a combination of the parameters from (23), different, thus, for (24) and (25). Comparing these combinations, (26) will have different formats and Teräsvirta (1994) elaborates successive F-tests which can distinguish between on function and another. They are[7]:

1. $H'_{01} : \beta_4 = 0$.
2. $H'_{02} : \beta_3 = 0 | \beta_4 = 0$.
3. $H'_{03} : \beta_2 = 0 | \beta_3 = \beta_4 = 0$.

If H'_{01} is rejected, we have evidence in favor of the LSTAR model, otherwise, we can use the ESTAR model. If H'_{10} and H'_{20} are not rejected, we also have evidence in favor of LSTAR, the hypothesis rejection, however, is not informative. If the true model is a LSTAR, H'_{03} is generally rejected. If H'_{03} is not rejected and H'_{02} is rejected, we choose the ESTAR model. If H'_{03} and H'_{02} are not rejected, but H'_{01} rejected, LSTAR model is selected. The only inconclusive case is when H'_{01} and H'_{02} are rejected. In this situation, we test:

$$H''_{02} : \beta_3 = 0 | \beta_2 = \beta_4 = 0. \qquad (28)$$

However, if H'_{02} is rejected, H''_{02} should be rejected even more strongly.

[7] The notation $\beta_i = 0 | \beta_j = 0$ means that the test $\beta_i = 0$ is made conditional on the non-rejection of the previous test $\beta_j = 0$. Thus, H_{02} is made conditional on the non-rejection of H_{01}, and H_{03} is made conditional on the non-rejection of H_{02} and of H_{01}.

Parameter estimation can be done by Nonlinear Least Squares, that, under the assumption of normality, is equivalent to the estimation of Conditional Maximum Likelihood. Under some assumptions and regularity conditions (Wooldridge, 1994), estimators are consistent and asymptotically normal. The estimation procedure is thus carried out with standard numerical algorithms (*e.g.* Newton-Raphson and BFGS).

This work's objective is to make one-step ahead forecasts, re-specifying the model at each period. It is, therefore, impracticable and counter-intuitive to decide, at each step, if the DGP comes from a STAR model with logistic function or with an exponential function. Impracticable for the fact that the test explained above for the most appropriate function contains inconclusive situations, making it difficult to construct an algorithm for real-time automatic forecast. Counter-intuitive because the change in the function at each step would imply that, given the information at each new observation, the entire series generating process would be revalued, being governed alternatively by a function characterizing asymmetric series and by a function characterizing symmetric series.

Since, usually, industrial production tends to have an asymmetric behavior, *i.e.*, its growth dynamics is different from its recession dynamics, we opted to use only the logistic function in the specification of the STAR model.

Castle and Hendry (2014) investigate, through Monte Carlo simulations, the LSTAR models forecast and estimation efficiency for different magnitudes in the transition probability from one regime to another and differences between the intercepts of each regime. They conclude that, for very high γ's, estimation becomes imprecise. Besides, there must be a large number of observations in each regime. For this to happen in small samples, it would be necessary a large probability of change between regimes. On the other side, the constant change of regimes can imply that the series is better modeled by a linear autoregressive process. These problems tend to be solved in reasonably large samples.

Concerning the forecast results, precision, measured by the Root Mean Squared Forecast Error (RMSFE), of a correctly specified LSTAR with one lag improves considerably in large samples and when the difference in the mean of each regime is not very large nor very small (the cited authors analyze changes of 1, 2, and 5 standard-deviations in the intercept).

From equation (23), the one-step ahead forecast, conditional to the information up to time t (I_t), is given by:

$$\begin{aligned} \mathrm{E}[y_{t+1}|I_t] &= \mathrm{E}[\pi_{10} + \pi_1^{\prime} z_{t+1} + (\pi_{20} + \pi_2^{\prime} z_{t+1}) F(y_{t+1-d}) + \ _{t+1}|I_t] \\ &= \pi_{10} + \pi_1^{\prime} z_{t+1} + (\pi_{20} + \pi_2^{\prime} z_{t+1}) F(y_{t+1-d}), \end{aligned} \tag{29}$$

where the terms z_{t+1} and $F(y_{t+1-d})$ are values already realized in period t, because $z_{t+1} = (y_t, y_{t-1}, \ldots, y_{t-p-1})$ and $d \geq 1$.

More-steps ahead forecasts are more complex, because obtaining $\mathrm{E}(y_{t+2}|I_t)$ is more difficult and would have to be done by numerical integration. This work, however, makes only one-step ahead predictions, which dismisses a deeper explanation of the subject[8].

[8]See Lundbergh and Teräsvirta (2005) for more detailed reference in forecasting more than one steps ahead in Smooth Transition Autoregressive models.

3.3 Autometrics

Autometrics has its origin in the so-called London School of Economics methodology, whose econometric technique is based on a general-to-specific modeling. In this approach, the researcher chooses a reasonably large number of variables which have a theoretical relation with some dependent variable and, from this general model, he performs specification and significance tests to gradually reduce the set of relevant variables.

This methodology opposes to the so called specific-to-general, that is more frequently used by econometrics practitioners. It consists in starting with simple models containing few variables and, according to diagnostic tests results, adding other variables or modifying the used specification.

The algorithm called Autometrics is presented in Doornik (2009). This type of algorithm, however, has its origin in the works of Lovell (1983) and Hoover and Perez (1999). Later, Hendry and Krolzig (1999, 2005) improve previous works, arriving at a structure similar to Autometrics, which has been continuously improved since then.

Hoover and Perez (1999) consider four main characteristics in their selection mechanism. First, the concept of General Unrestricted Model (GUM), that is a linear regression model involving a large number of variables relating with the utilized dependent variable. Second, the use of multiples search paths, where the first path begins with the elimination of the variable with the largest p-value in the significance test, the second path with the elimination of the variable with the second largest p-value and so on until the tenth variable. Third, at each step of each path, it is performed a joint significance test including all the eliminated variables so far in relation to the GUM. Fourth, at each step of each path, a battery of specification tests are also performed.

In Hendry and Krolzig (1999, 2005), the authors improve the existing algorithm, creating the software called *PcGets* (from general-to-specific). Differently for the previous mechanism, more paths are followed eliminating sets of variables, and there is a pre-selection of lags before the reduction begins. Besides, each path's "terminal" — when it is not possible to remove any more variable without failing in some test — are united in a set of variables which will form the new GUM. The algorithm is finished when there is convergence, *i.e.*, the current GUM equals the previous GUM.

In the end, we will have many candidates which cannot be reduced. Among them, the final candidate is chosen by the minimum Bayesian information criterion (BIC) in Hendry and Krolzig (1999, 2005) and by the best fitted model in Hoover and Perez (1999).

3.3.1 Autometrics' Selection Algorithm

In Doornik (2009), starting with the GUM, the selection mechanism seeks to consider the models with all possible variable combinations. It utilizes a *tree representation* (Figure 1), in which, each letter represents one variable and each node a model to be estimated.

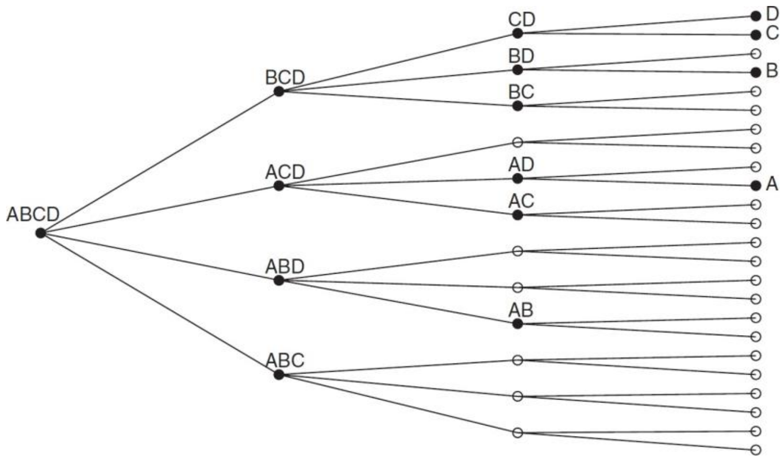


Figure 1: Tree for Possible Models in Autometrics.

The algorithm runs through the tree, following the nodes from left to right and from top to bottom, so that, in the Figure 1 example, it would pass through the models ABCD, BCD, CD, D, C, BD and so on. In each node, the variables are ordered in crescent order of significance, that is, in the GUM, A is the least significant variable and D is the most significant. It is possible that, in the following node, the order will be changed. For example, if C is less significant than B, the ordering would be given by CBD and the variable to be eliminated would be C.

If all the nodes from the tree were evaluated, the computational cost would be very large (for n regressors, $2^n - 1$ models would be estimated), so a number of strategies to jump the irrelevant nodes are used. These strategies are classified by Doornik (2009) as Pruning, Bunching, Chopping and Model Contrast.

Pruning consists in ignoring the nodes derived from a node invalidated by a diagnostic test or by backtesting (a joint significance F-test with relation to the initial GUM). If, *e.g.*, starting from BCD, variable B cannot be eliminated, models CD, D and C are not estimated. The p-value used is denoted p_a.

Bunching stems from the idea that it is possible, instead of eliminating the variables individually based in a t-test, eliminating groups of variables based in a F-test with relation to the current node. The size of the group to be tested is defined by $p_b = max\{\frac{1}{2}p_a^{1/2}, p_a^{3/4}\}$[9], the p-value of the joint significance test. The F-tests significance level for which the group is eliminated, however, is the same as Pruning's, *i.e.*, p_a. Once the group of variables is eliminated, we get to the correspondent following node, avoiding estimation of many models. In the example of Figure 1, if we are at BCD and are able to effectively eliminate jointly B and C, the next model to be estimated is that containing only D, ignoring CD. If, in group, the variables cannot be withdrawn, we test the removal of groups with less variables, until a valid reduction is achieved.

Chopping means ignoring models, in a given ramification from a node, containing some variable or group of variables if they are of very little significance. For example, if we are, again, in BCD and BC presents a F-test with a greater p-value than a certain p_c, the algorithm

[9] See Doornik (2009) for an explanation about this value.

will estimate, in this ramification of the tree, only model D and will go next to node ACD. The standard p-value used by Autometrics is $p_c = p_b$.

Model Contrasts consist in the elimination or shortening of paths in a certain ramification of the tree if it contains some terminal already found previously, so that it jumps to the nodes possessing possible terminals not yet estimated. As an example, suppose that, in the above Figure, we had selected D as a terminal in the first ramification. In the next step, ACD, the following node, AD, would originate two possible models: A and D. As D was already selected, the algorithm tests directly a reduction of ACD to A, speeding the calculations.

There are two types of Model Contrast: Union Contrast and Terminal Contrast. In the first, used when the current GUM is different from the previous GUM, the program considers, to jump, nodes that lead to any model different from the terminal union. In the second, used when the current GUM equals the previous, the jumped nodes takes into account paths that lead to terminals different from each one of the previously individually found terminals.

Several specification tests are then performed. However, differently from the algorithms cited earlier, in Autometrics they are performed only when the terminals are attained. If some test fail in some terminal, the program follows the path backwards from the node, making specification tests until some valid terminal is found. The tests performed are: residuals normality (Jarque and Bera, 1980), second order residual autocorrelation (χ^2 test, Godfrey, 1978; Breusch and Pagan, 1980), autocorrelated conditional heteroscedasticity (ARCH) to second order (Engle, 1982) and that of in-sample stability (Chow, 1960).

In each iteration of the GUM, Autometrics divides the search in two stages. In the first, it ignores paths whose nodes contain terminals. In the second, it follows the root paths which contain terminals, using Model Contrast to reach definitive terminals more quickly.

Furthermore, before initializing the algorithm *per se*, there is the option of performing a presearch to eliminate sets of variables or lags. The general model after the presearch is used as the GUM in the algorithm. Several options of lag and variable reduction are available. It is out of the scope of this paper to explain the details of the presearch procedures. For a description of the different options, see Doornik (2009).

Finally, when all the tree paths are covered, the final terminals where specification tests failed are discarded and, after the GUM convergence, *i.e.*, when the previous GUM is equal to the current GUM, the terminal with the minimum BIC is chosen.

3.3.2 Dummy Saturation

A way to detect discrepant observations (outliers) in the series and other possible nonlinearities is the use of impulse dummy variables — which assume value 1 for observation t and 0 for the others — for the periods they occur. Proceeding according to the *Gets* methodology, we would have one dummy variable for each observation, besides other exogenous variables which could possibly affect the dependent variable. This framework is denoted as a Dummy Saturation model.

Trying to estimate the GUM this way is obviously impracticable, because there are more variables than observations. Autometrics, however, utilizes a technique to reduct these models, initially proposed by Hendry and Krolzig (2005) and applied to the Impulse Indicator Saturation (IIS) context by Santos et al. (2008).

In case no variables beside the dummies are included, the first step is to apply the reduc-

tion mechanism described above only with the T/2 first dummies, so that the resulting model will contain a subset of them. After this, the same is done with only the T/2 last dummies. Thus, we will obtain two models containing subsets of the initial dummies, the union of these models will generate a new GUM. Applying once more the reduction mechanism, we obtain the model containing the relevant dummies.

A generalization of this method can be applied when step dummies — which assume value 0 up to observation t_1 and 1 in the following observations — and other exogenous variables are included in the GUM. In this case, the algorithm introduced by Hendry and Johansen (2012) assumes $N = \sum_{j=1}^{N} n_j$ regressors so that $N > T$ and $n_j < T$ for all j, that is, the N regressors are partitioned in portions smaller than the sample size.

Once the best fitted model is found, with relevant variables and dummies, the forecast is made in the standard way, *i.e.*, the model's n-steps ahead mathematical expectation.

3.4 Double Difference Device

Consider a variable y_t that we want to predict[10]. Suppose $y_t \sim D_{y_t}(y_t|Y_{t-1}, \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^k$ and $Y_{t-1} = (y_1, \ldots, y_{t-1})$. For a sample of size T, a forecast h step ahead is a combination of the sample values from 1 through T, *i.e.*, $\hat{y}_{T+h|T} = f_h(Y_T)$. It can be proven (for example in Clements and Hendry, 1998) that $\bar{y}_{T+h|T} = \mathsf{E}[y_{T+h}|Y_T]$ is the unbiased predictor which minimizes the Mean Square Forecast Error (MSFE) and holds the minimum variance within the unbiased predictor class.

Suppose y_t follows the linear autoregressive process below:

$$y_t = \mu + \rho y_{t-1} + \gamma z_t + \epsilon_t, \text{ with } \epsilon_t \sim IN(0, \sigma^2_\epsilon) \text{ e } |\rho| < 1, \tag{30}$$

where $\{z_t\}$ is an exogenous variable, forecast error's mean and variance are, respectively, $\mathsf{E}[y_{T+1} - \bar{y}_{T+1|T}] = 0$ and $V[(y_{T+1} - \bar{y}_{T+1|T})] = \sigma^2_\epsilon$.

Consider $\mathsf{E}[y_t] = \theta$ and $z_t \sim N(\kappa, 1)$. We can write the mean of y_t as:

$$\mathsf{E}[y_t] = \theta = \frac{\mu + \gamma\kappa}{1 - \rho}. \tag{31}$$

If $\mu = \kappa = 0$, a break in ρ will not affect the mean of y_t. However, in case that we have one of these parameters different from zero, a change in the autoregressive parameter will imply a change in the mean. Besides this, if more than one of these parameters suffer a break, the forecast failure can be even more severe. For example, if the autoregressive parameter and the intercept, ρ and μ, are, respectively, 0.8 and 8, and suffer a break, becoming 0.6 and 6, the unconditional expectation shifts from $\theta = 40$ to $\theta^* = 15$.

Let us rewrite the model (30) as:

[10] This exposition is based in Hendry (2012), and simplifies the more general theory with multiple variables in Hendry (2006).

$$\Delta y_t = \mu + (\rho - 1)y_{t-1} + \gamma z_t + {}_t + (\rho - 1)\frac{\mu + \gamma\kappa}{1 - \rho} - (\rho - 1)\frac{\mu + \gamma\kappa}{1 - \rho}$$

$$= \mu - \mu - \gamma\kappa + (\rho - 1)(y_{t-1} - \theta) + \gamma z_t + {}_t$$
$$= (\rho - 1)(y_{t-1} - \theta) + \gamma(z_t - \kappa) + {}_t.$$

Rearranging the terms, i steps ahead values can be written as:

$$\Delta y_{T+i} = \omega + \alpha(y_{T+i-1} - \theta) + \gamma z_{T+i} + {}_{T+i}, \tag{32}$$

where $\omega = -\gamma\kappa$ and $\alpha = \rho - 1$.

This formulation is denoted Equilibrium Correction Model (EqCM), where the equilibrium is given by θ. Forecasts will tend to return to θ independently of the behavior of the data. Changes in this equilibrium, which occur mainly by shifts in the level, consist in the main factor which imply forecast failure.

A naive predictor is the called Double Difference Device (DDD). Its intuition comes from the fact that most economic variables do not continuously accelerate and, hence, its second differences have unconditional expectation $\mathsf{E}[\Delta^2 y_t] = 0$, suggesting the following DDD estimator:

$$\Delta y_{T+1|T} = \Delta y_T, \tag{33}$$

where $\Delta y_{T+1|T}$ denotes a one-step ahead forecast for Δy_T given the information in time T.

Because the forecast does not utilizes any parameter, structural breaks do not persistently influence it, and the estimator is unbiased.

Let us modify equation (32) so that the DGP does not contain exogenous variables besides the dependent variable lags (*i.e.* $\gamma = 0$):

$$\Delta y_{T+i} = \alpha(y_{T+i-1} - \theta) + {}_{T+i}. \tag{34}$$

Suppose there is a break in the parameters, so that the DGP becomes:

$$\Delta y_{T+i} = \alpha^*(y_{T+i-1} - \theta^*) + {}_{T+i}. \tag{35}$$

EqCM forecast error is given by:

$$\Delta y_{T+i} - \Delta y_{T+i|T+i-1} = \Delta y_{T+i} - \hat{\alpha}(y_{T+i-1} - \hat{\theta}) + {}_{T+i} = w_{T+i}, \tag{36}$$

where the hat over the parameters means that they were estimated on the EqCM formulation.

Replacing the in-sample estimated values for the pseudo-true in-sample population values, where $\mathsf{E}[\hat{\theta}] = \theta_p$, we can reduce the forecast error variance without damage to the general analysis. Thus, we have:

$$\mathsf{E}[w_{T+i}|y_{T+i-1}] = (\alpha^*\theta^* - \alpha_p\theta_p) + (\alpha^* - \alpha_p)y_{T+i-1}$$

$$V_{T+i}[w_{T+i}|y_{T+i-1}] = \sigma_E^2. \tag{37}$$

With respect to the DDD ($\Delta y_{T+i|T+i-1} = \Delta y_{T+i-1}$), for $i > 1$, we have:

$$\Delta y_{T+i-1} = \alpha^*(y_{T+i-2} - \theta^*) + {}_{T+i-1}. \tag{38}$$

The respective forecast error is calculated as:

$$\begin{aligned}\Delta y_{T+i} - \Delta y_{T+i|T+i-1} = u_{T+i} = \alpha^*(y_{T+i-1} - \theta^*) + {}_{T+i} \\ - [\alpha^*(y_{T+i-2} - \theta^*) + {}_{T+i-1}]\end{aligned} \tag{39}$$

$$\Rightarrow u_{T+i} = \alpha^* \Delta y_{T+i-1} + \Delta {}_{T+i}. \tag{40}$$

On the long term, values are replaced by their unconditional expectations, *i.e.*:

$$\mathsf{E}[u_{T+i}] = \alpha^* \mathsf{E}[\Delta y_{T+i-1}] + \mathsf{E}[\Delta {}_{T+i}] = 0 \tag{41}$$

$$\begin{aligned} V[u_{T+i}] &= V[\alpha^* \Delta y_{T+i-1}] + V[\Delta {}_{T+i}] \\ &= \alpha^{*2} V[\Delta y_{T+i-1}] + 2\sigma_E^2. \end{aligned} \tag{42}$$

As it is evident from (40) and (42), the Double Differencing mechanism adds some noise sources by the extra differentiation of $\alpha^* y_{T-i+1}$ and of ${}_{T+i}$. This extra source of noise, however, can be of lower dimension than those from the traditional autoregressive model when there is presence of structural breaks in the series' unconditional mean.

To illustrate, suppose that the shift occurs only in μ, so that θ changes and α remains constant. We will have:

$$\begin{aligned} w_{T+i} &= -\alpha(\theta^* - \theta) + {}_{T+i} \\ \mathsf{E}[w_{T+i}] &= -\alpha(\theta^* - \theta) \\ V[w_{T+i}] &= \sigma_E^2 \\ u_{T+i} &= \alpha \Delta y_{T+i-1} + \Delta {}_{T+i} \\ \mathsf{E}[u_{T+i}] &= 0 \\ V[u_{T+i}] &= V[\alpha \Delta y_{T+i-1}] + V[\Delta {}_{T+i}]. \end{aligned}$$

The MSFE is approximately:

$$M[w_{T+i}] = \alpha^2(\theta^* - \theta)^2 + \sigma_E^2. \tag{43}$$

Comparing with the DDD MSFE:

$$M[u_{T+i}] = 2\sigma_E^2 \quad 1 + \frac{\alpha^2}{2+\alpha} \quad . \tag{44}$$

Assuming $\rho = 0.8$, $\nabla \mu^* = \mu^* - \mu = 0.2$, and $\sigma_E = 0.06$, we have $\alpha = -0.2$ and $\nabla \theta^* = \theta^* - \theta = 1$. With these values, $M[w_{T+i}]$ is approximately 6-fold larger than $M[u_{T+i}]$.

3.5 LASSO and Derived Methods

Developed by Tibshirani (1996), the Least Absolute Shrinkage and Selection Operator (LASSO) is based on a class of estimation procedures known as shrinkage methods, which shrink Ordinary Least Squares (OLS) coefficients estimators towards zero. This approach aims to reduce the variance in the bias-variance trade-off, introducing some bias in the estimators.

Consider the data $(\mathbf{x^i}, y_i)$, $i = 1, \ldots, N$, where $\mathbf{x^i} = x_{i1}, x_{i2}, \ldots, x_{ip}$ are the regressors and y_i the dependent variable, and assume that the y_is are conditionally independent given the x_{ij}s. The LASSO estimator solves the following problem:

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \Big(y_i - \alpha - \sum_j \beta_j x_{ij} \Big)^2, \text{ subject to } \sum_j |\beta_j| \le t, \tag{45}$$

where $t \geq 0$ is a tuning parameter. The independent variables are standardized (*i.e.*, $\sum_i x_{ij}/N = 0$ and $\sum_i x_{ij}^2/N = 1$) to rule out problems of differences in their scales. For all t, the solution for α is given by $\hat{\alpha} = y$, so that (45) can be solved omitting the intercept.

The problem can be expressed in its equivalent Lagrangian form:

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \Big(y_i - \alpha - \sum_j \beta_j x_{ij} \Big)^2 + \lambda \sum_j |\beta_j|. \tag{46}$$

For $\lambda = 0$ (for a sufficiently large t), the restriction in (46) is nonbinding and $\hat{\beta}_{LASSO} = \hat{\beta}$, the OLS estimator. Higher values of λ impose the estimators a penalty, shrinking them towards zero. Depending on the value of λ, some of the LASSO coefficients are set to be exactly equal to zero, performing, thus, a method of variable selection.

To better understand this characteristic of the LASSO, consider an earlier example of regularization method introduced by Hoerl and Kennard (1970), the Ridge Regression estimator, which solves the following problem (also in its Lagrangian form):

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \Big(y_i - \alpha - \sum_j \beta_j x_{ij} \Big)^2 + \lambda \sum_j \beta_j^2, \tag{47}$$

with the same set-up of the LASSO.

The difference between the two methods lies in the form of the penalty: the LASSO uses the L^1-norm penalty and the Ridge Regression uses the L^2-norm penalty.

Due to the L^1-norm used in LASSO, the solution to (46) is nonlinear in y_i, for which there is no closed form expression. For the Ridge Regression, the solution takes the form $\hat{\beta}_{RIDGE} = (X'X + I\lambda)^{-1}X'y$. Note that, differently from the LASSO, $\hat{\beta}_{RIDGE}$ will generally have all its components different from zero. To see this, consider Figure 2 below.

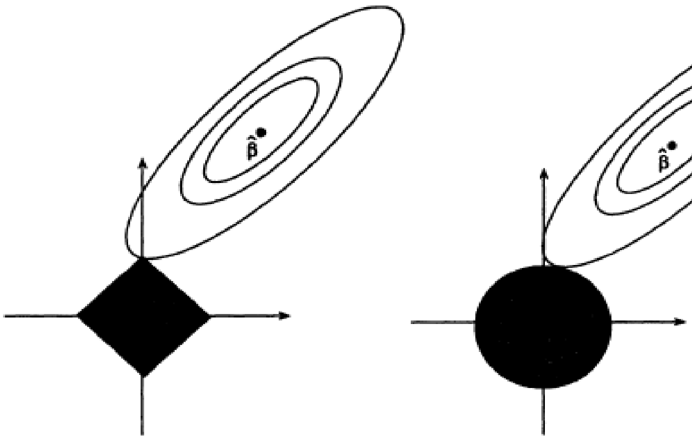


Figure 2: Estimation picture for LASSO (left) and Ridge Regression (right). Source: Tibshirani (1996).

In a two parameter space, the OLS estimator is represented by $\hat{\beta}$, the ellipses are contour plots for the same sum of squared residuals, and the black area represents the possible values under the constraint for some t. The corners on the restriction area make the LASSO estimator, in the left, to fix one of the coefficients exactly at zero. The Ridge Regression estimator, in the right, shrinks the OLS estimator towards zero, but does not, as a rule, estimate exactly zero coefficients.

Up to this point, our analysis treated λ as given. In practice, we have to choose it according to the data. Probably, the most common approach to estimate λ is by cross-validation, which aims to minimize an estimate of the expected prediction error[11]. However, cross-validation may present difficulties in a time-series framework. Zou et al. (2007), Zhang et al. (2010), Wang et al. (2007) argue that using the Bayesian Information Criterion (BIC) is a good alternative to cross-validation in selecting λ. We follow Medeiros and Mendes (2016), Medeiros and Vasconcelos (2016) and Konzen and Ziegelmann (2016), for example, in using the BIC to select the tuning parameter.

3.5.1 AdaLASSO

Fan and Li (2001) defined a procedure satisfying the oracle properties as one that, asymptotically, selects the correct subset of variables with nonzero coefficients and has an optimal estimation rate. Zou (2006) states that if, for some λ, the LASSO has an optimal estimation rate, then it does not satisfy variable selection consistency. The author also argues that, even relaxing an optimal estimation rate, variable selection is consistent only under a nontrivial condition (similar conclusions were made by Meinshausen and Bühlmann, 2006 and Zhao and Yu, 2006).

Zou (2006) proposed, thus, a slight modification to the LASSO, the Adaptive LASSO (AdaLASSO) estimator, which solves the following problem:

[11] See Hastie et al. (2009) for a more detailed explanation of the cross-validation method for model selection and parameter tuning.

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 + \lambda \sum_j \hat{w}_j |\beta_j|, \tag{48}$$

where $\hat{w}_j = |\hat{\beta}_{FSTEP}|^{-\gamma}$, $\gamma > 0$, and $\hat{\beta}_{FSTEP}$ is a first step estimator of β. One can use, for example, the OLS estimator, the Ridge Regression estimator, or the LASSO estimator (as in Garcia et al. (2017)). We opt to use $\hat{\beta}_{FSTEP} = \hat{\beta}_{RIDGE}$ and $\gamma = 1$.

The AdaLASSO assigns different penalties to each variable, with higher penalties to variables whose coefficients are closer to zero in the first step estimation. Zou (2006) proves that the AdaLASSO is an oracle procedure.

3.5.2 WLAdaLASSO

Another variant of the LASSO we use is the Weighted Lag Adaptive LASSO (WLAdaLASSO), proposed by Konzen and Ziegelmann (2016), based on the work by Park and Sakaori (2013). This method is similar to the AdaLASSO, and it comes from the observation that, in a time-series autoregression framework, more distant lags tend to have less influence in forecasting the dependent variable, imposing on them higher penalties. The estimator solves:

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 + \lambda \sum_j \hat{w}_j^{WL} |\beta_j|, \tag{49}$$

where $\hat{w}_j^{WL} = (|\hat{\beta}_{RIDGE}| e^{-\alpha l})^{-\gamma}$, l is the lag order, and $\gamma > 0$, $\alpha \geq 0$ are tuning parameters. As in AdaLASSO, we set $\gamma = 1$. To select α, for a given λ, we estimate the model considering a grid $(0, 0.5, 1, \ldots, 10)$ for α and choose that with the lowest BIC. The parameter λ is then selected considering the model with the lowest BIC (see Konzen and Ziegelmann, 2016 and Prince and Marçal, 2018).

Monte Carlo simulations performed by Konzen and Ziegelmann (2016) pointed that the WLAdaLASSO was superior to both LASSO and AdaLASSO in variable selection, parameter estimation and forecasting, particularly when the candidate variables included a high number of lags and predictors presented stronger linear dependence.

4 Empirical Framework

The objective of this work is twofold. The first is to test if there is any gain in using extended series of Brazilian Industrial Production, the second is to assess if nonlinear univariate methods have better forecasting performance than the autoregressive model.

This section details the PIM-PF time-series available and describes the forecasting exercise we perform. It also details the method to test whether forecasts accuracy are statistically different between models.

4.1 PIM-PF Data

There are, as mentioned in the previous paper, three different series available in IBGE's System of Automatic Data Retrieval (SIDRA): 1985:01-2004:01, with base year 1991, 1991:01-2014:02, with base year 2002, and 2002:01-2018:12, with base year 2012. There is an extended version of the first series available in the Time Series Management System (SGS) of Brazilian Central Bank (BCB), ranging from 1975:01 to 2004:01, where the 1985:01-2004:01 values are identical to those from SIDRA.

Figure 3 plots the three series in one graph. Note that there is considerable overlap between their covered time span. This is because the closed series published by IBGE already contain chained values from the previous one.

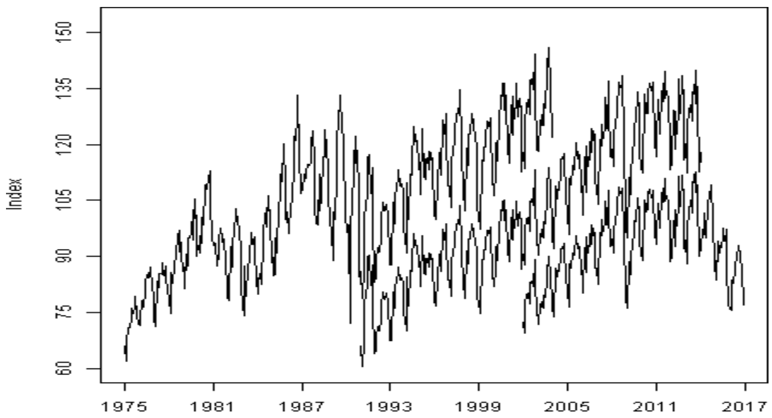


Figure 3: PIM-PF Series Expressed in Different Base Years.

One reason the entire chained series is not always published is that there are difficulties in the chaining components disaggregated by activities, due to changes in the classification system and different list of selected products in each revision. The chaining of the aggregated series is, on the other hand, straightforward and plausible. Figure 4 contains the chained series, constructed as described in (3) and (5), of the General Industrial Production *quantum* for whole sample.

The presence of non-stationarity and seasonality in the series is easily noticed. To deal with these characteristics, the log of the series is differenced with relation to the same month in the previous year. This corrects both problems if the series is integrated of order one and has constant seasonality[12]. Figure 5 shows a graph with the treated series.

For nonlinear models with different regimes, specification becomes more robust when data shows well defined different regimes. In STAR models, for example, γ estimation becomes difficult if the probability of crossing the threshold c is large, *i.e.*, the regime are not very well

[12]The presence of stochastic seasonality, however, may not be eliminated with this transformation. In fact, even standard seasonal adjustment methods may fail to correctly account for stochastic seasonality. The use of Autometrics with dummy saturation, however, can improve the modelling of non-deterministic seasonality.

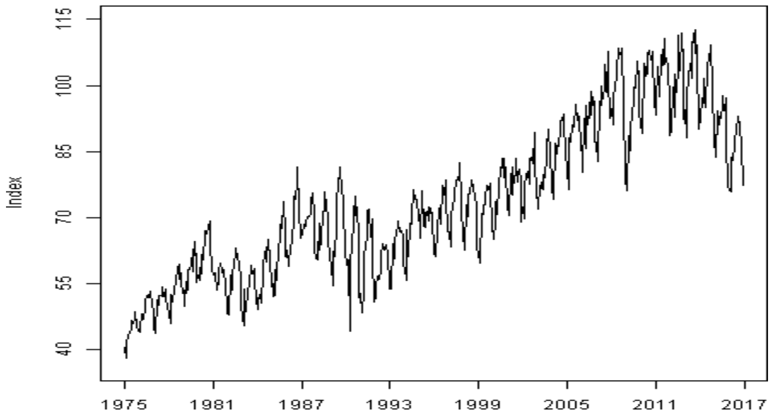


Figure 4: Chained PIM-PF Series. Mean 2012=100

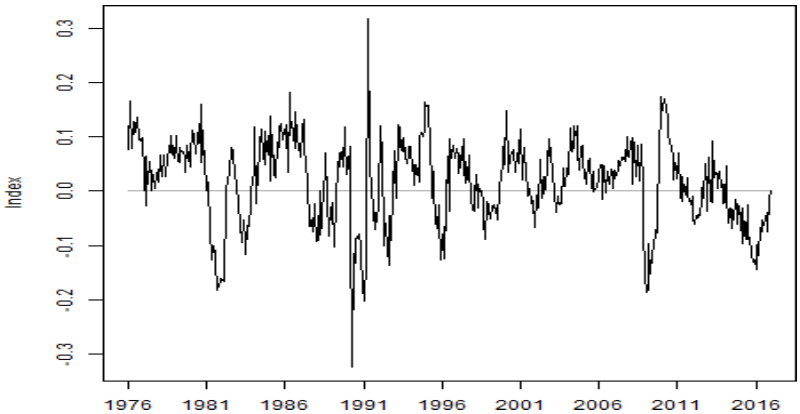


Figure 5: Log General Industrial Production - Annual Difference

defined (Castle and Hendry, 2014; Granger and Teräsvirta, 1993). The Markov Switching models also adapt better when there are well defined regimes.

4.2 Description of the Forecasting Exercise

This work evaluates the performance of forecasts using samples of different lengths from the Brazilian Industrial Production Index. Given the volatile and asymmetric nature of the investigated variable, it is also of interest to compare forecasts accrued from nonlinear methods with those of a standard benchmark methods. Therefore, we make two forecasting exercises. One compares, for each method, forecasts using chained series, extended to different starting periods, with forecasts using the latest PIM-PF series available, beginning in 2002:01. The other compares, for each sample, forecasts of different methods with those of AR(p).

Forecasts are made one-step ahead, re-specifying and re-estimating the models at each step. The AR and STAR models order p is selected by the BIC. To specify the MSAR model, at each forecast step, we estimate 27 different models (1 to 13 lags with switching autoregressive parameters, 1 to 13 lags with fixed autoregressive parameters, and one only with intercept, all with two regimes and including switching intercept and variance) selecting that with the lowest BIC. LASSO, AdaLASSO, WLAdaLASSO and Autometrics algorithms are reapplied at every step. For all methods, a maximum lag order of 13 is considered. In Autometrics, we use a standard p-value of 0.01 for significance tests and saturations of impulse and step indicators.

Since we used a twelve month difference to treat the data, we will suppose $E[\Delta\Delta_{12}y_t] = 0$ instead of $E[\Delta^2 y_t] = 0$ in the double difference device and denote it DD12, *i.e.*, the industrial production growth with respect to the same month of the previous year do not continuously accelerate. This modication is straightforward and the interpretation is not impaired.

We analyze the forecast horizon starting in January 2007 and ending in December 2016, comprising a total of 120 months, in which forecasts are made one-step ahead. We also analyze the forecast performance in both halves of the forecast horizon, *i.e.*, 2007:01-2011:12 and 2012:01-2016:12, in which the first half is characterized by abrupt changes in the series, reflecting the effects of the subprime crisis in Brazilian industrial production and its subsequent recovery, and the second by a less abrupt, but persistent, decline in the industrial activity level.

Results are obtained using the chained series starting in 1975:01, 1985:01 and 1991:01. We also consider forecasts made with a 121 months rolling window and with an extending window with the same initial size, *i.e.*, starting in 1993:12[13]. Note that these starting periods coincide with those from the closed series obtained from SIDRA and BCB. Using these samples may help identify if more distant methodological benchmarks contain useful information for forecasting recent values.

4.3 Forecast Comparison (Diebold and Mariano Test)

To compare the forecasts, we used the Diebold and Mariano (1995) test, which utilizes the forecast errors from two different models to evaluate if one forecast is statistically superior the other.

For the test to be valid, it suffices that assumptions about the forecast errors are valid and it is not necessary to make hypotheses about the models being tested, so that it is possible to compare even predictions that do not come from models.

Let e_{it} be the forecast error from model i in period t and $g(e_{it})$ some loss function. In this study, we analyze the results for both $g(e_{it}) = e^2_{it}$, the quadratic loss function, and $g(e_{it}) = |e_{it}|$, the absolute loss function. The key hypotheses are actually made about the difference between the loss function associated with the forecast errors, $d_t = g(e_{it}) - g(e_{jt})$. It is assumed that d_t is covariance-stationary:

[13]These are the starting months of the level series, the annual difference of the index logarithm drops 12 observations from each sample.

$$E(d_t) = \mu, \forall t$$
$$cov(d_t, d_{t-\tau}) = \phi(\tau), \forall t$$
$$V(d) < \infty$$

The hypothesis of equal predictive capacity is equivalent to $E(d_t) = 0$. In this case, the test statistics is:

$$DM = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}} \xrightarrow{d} N(0,1) \tag{50}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}(g(e_{it}) - g(e_{jt}))$ and $\hat{\sigma}_{\bar{d}}$ is a consistent estimator for the standard deviation of d.

As d_t has frequently some autocorrelation, mainly due to imperfect predictions, DM statistic must be calculated with a robust $\hat{\sigma}_{\bar{d}}$. Diebold and Mariano (1995) suggest $\hat{\sigma}_{\bar{d}} = 2\pi\hat{f}_d(0)/T$, where $\hat{f}_d(0)$ is a consistent estimator for the spectrum of the loss function differential at frequency zero. $f_d(0) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\phi(\tau)$.

One caveat must be made concerning this method. Its objective is to test if different predictions from diverse sources are statistically different and, for this, it uses as primitive the forecast error loss function differential. When we compare two models, saving a fraction of the sample to calculate the loss function, what is used as primitives for the Diebold and Mariano test are, ultimately, the estimated parameters of the model, which originate the forecasts and, thus, the forecast errors. In doing so, part of the sample is lost, increasing the estimation uncertainty and jeopardizing efficiency in small samples.

Diebold (2015) argues that there are consolidated methods using the whole sample which are more efficient in model comparison. In a Bayesian context, the best model selection criterion resumes to that with the largest marginal likelihood function, that, as shown in Schwarz (1978), corresponds asymptotically to the model with smallest BIC. In fact, in the Gaussian linear regression context, the Bayesian information criterion can be written as $BIC = T^{(k/T)}MSFE$, where $MSFE = \frac{\sum^{T} e^2}{T}$, that is, the BIC is an estimate of the out-of-sample MSFE.

Comparing different models, however, is not the only focus of this work. Other two important objectives are: (i) test if using longer samples, with series produced using different methodological benchmarks, is superior to using only the last available sample; (ii) assess how one-step ahead forecasts for different methods behave in different periods and different estimation samples. Diebold (2015) criticizes the indiscriminate use of the Diebold and Mariano test to compare models, but argues that it may be useful to, among other things, flag for structural breaks and to assess predictive performance over different business cycles. Furthermore, at each forecast step a new model is specified and estimated, and estimation using different sample sizes are performed. Therefore, Diebold and Mariano test arrive as an appropriate test, since it uses only the series of loss function differentials between two models.

5 Results

Tables 1 and 2 present the results of the Root Mean Squared Forecast Error (RMSFE) and Mean Absolute Forecast Error (MAFE) for each sample and model. We also present, in Table 3, the Mean Forecast Error (MFE), which gives an idea about the bias of forecasts.

Table 1: Root Mean Squared Forecast Error

2007:01 - 2016:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	4.08	4.70	4.25	4.20	4.18	4.28	4.17*	4.29	4.57	4.37	**4.01**
121-RW	3.88	4.31	4.08	4.02	3.89	3.89	4.17*	3.95	**3.77**	4.09	**3.77**
1993:12	3.76*	3.96*	3.96*	3.82	3.79	3.79	4.17*	3.81	3.83	3.82*	**3.65***
1991:01	3.88	4.35	3.97	3.85	3.76	3.76	4.17*	3.72	3.84	3.88	**3.71**
1985:01	3.99	4.27	4.05	3.94	3.80	3.87	4.17*	3.74	**3.60**	4.01	3.76
1975:01	3.93	4.04	3.97	3.81*	3.75*	3.75*	4.17*	3.64*	**3.57***	4.23	3.72
2007:01 - 2011:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	4.60	5.46	4.64	4.60	4.65	4.75	**4.30***	4.92	5.44	5.02	4.46
121-RW	4.43	4.70	4.55	4.37	4.26	4.26	4.30*	4.34	4.25	4.58	**4.13**
1993:12	4.25*	4.49	4.32	4.13	4.14	4.14	4.30*	4.32	4.45	4.20*	**4.02***
1991:01	4.38	5.10	4.30	4.15	**4.08**	**4.08**	4.30*	4.11	4.45	4.25	4.09
1985:01	4.37	4.65	4.38	4.12	3.99	4.05	4.30*	**3.95***	4.12*	4.54	4.05
1975:01	4.29	4.45*	4.27*	4.05*	**3.98***	**3.98***	4.30*	4.04	4.12*	4.77	4.04
2012:01 - 2016:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	**3.48**	3.79	3.81	3.75	3.65	3.75	4.03*	3.55	3.50	3.61	3.50
121-RW	3.24	3.89	3.54*	3.63	3.49	3.48	4.03*	3.51	**3.23**	3.54	3.38
1993:12	3.20*	3.34*	3.56	3.49*	3.41	3.41	4.03*	3.22	**3.08**	3.40	3.24*
1991:01	3.30	3.44	3.60	3.51	3.40*	3.40*	4.03*	3.29	**3.12**	3.47	3.30
1985:01	3.58	3.84	3.69	3.74	3.61	3.67	4.03*	3.53	**3.00**	3.39*	3.45
1975:01	3.53	3.58	3.63	3.56	3.50	3.50	4.03*	3.18*	**2.92***	3.62	3.35

Notes: Values are multiplied by 100. The first column denotes the initial observation of the estimation sample, 121-RW denotes a 121 months rolling window forecasting scheme. Lowest absolute values for each method (column) are marked by *, boldface numbers mark lowest absolute values for each sample (row).

Figure 6 plots RMSFE, MAFE and MFE for all models in the first, second and third lines, respectively. Results are shown for the whole forecast horizon (first column) and both of its halves (second and third columns). The dashed line links the median of plotted values (*i.e.*, median of RMSFE, MAFE and MFE values). Point forecast results and the value of RMSFE, MAFE and MFE medians can be found in the appendix.

We first focus on the difference between the results for different forecast periods. A first fact we may note is that forecasts for 2012:01-2016:12 are more accurate than for 2007:01-2011:12. In fact, the highest RMSFE for all models and samples considering only the second half of the forecast horizon, that corresponding to the double difference device (DD12), is only slightly higher than the lowest RMSFE for all models and samples considering only the first half, that of Autometrics selection without any dummy saturation (AUTO) with the sample starting in 1985:01 (4.05 for DD12 and 3.95 for AUTO).

The MAFE results are less discrepant between the different forecast horizons, although the median of MAFEs (the dashed lines in Figure 6) are also lower in the second half of forecast horizon for each sample. This arises from the fact that squared forecast error loss function punishes more heavily forecasts which are more distant from the actual realization

Table 2: Mean Absolute Forecast Error

2007:01 - 2016:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	**2.93**	3.29	3.27	3.15	3.05	3.16	3.22*	3.14	3.26	3.43	3.01
121-RW	**2.88**	3.19	3.02	3.13	3.06	3.04	3.22*	3.07	2.92	3.06	2.91
1993:12	**2.75***	2.86*	2.99*	3.02	3.00	3.00	3.22*	2.93	2.90	2.91*	2.79*
1991:01	2.87	3.02	3.04	2.98*	2.90*	2.90*	3.22*	2.88	2.85	3.03	**2.83**
1985:01	3.09	3.25	3.13	3.11	3.04	3.08	3.22*	2.95	**2.74**	3.09	2.92
1975:01	3.00	3.05	3.05	3.00	2.96	2.96	3.22*	2.86*	**2.67***	3.28	2.88

2007:01 - 2011:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	**3.10**	3.56	3.40	3.16	3.12	3.25	3.28*	3.39	3.78	3.88	3.14
121-RW	3.14	3.38	3.16	3.20	3.16	3.16	3.28*	3.21	3.28	3.27	**2.99**
1993:12	2.90*	2.98*	3.08*	3.06	3.09	3.09	3.28*	3.15	3.21	2.94*	**2.83***
1991:01	3.01	3.22	3.14	2.97*	2.91*	2.91*	3.28*	2.95*	3.10	3.15	**2.88**
1985:01	3.30	3.37	3.23	3.08	3.03	3.07	3.28*	2.97	**2.93**	3.32	2.96
1975:01	3.16	3.22	3.14	3.00	2.97	2.97	3.28*	3.03	**2.89***	3.57	2.96

2012:01 - 2016:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	2.77	3.02	3.13	3.13	2.99	3.06	3.16*	2.90	**2.74**	2.99	2.88
121-RW	2.63	3.00	2.88*	3.06	2.96	2.91	3.16*	2.92	**2.56**	2.85	2.82
1993:12	2.59*	2.73*	2.90	2.98*	2.91	2.91	3.16*	2.72	**2.58**	2.88	2.74
1991:01	2.73	2.82	2.95	3.00	2.88*	2.88*	3.16*	2.80	**2.59**	2.91	2.77*
1985:01	2.89	3.12	3.03	3.13	3.05	3.09	3.16*	2.93	**2.54**	2.85*	2.87
1975:01	2.83	2.89	2.97	3.00	2.95	2.95	3.16*	2.68*	**2.45***	3.00	2.80

Notes: see notes from Table 1.

Table 3: Mean Forecast Error

2007:01 - 2016:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.22	0.85	0.43	1.21	1.10	1.02	**0.00***	0.32	0.67	0.62	0.64
121-RW	0.16	0.75	0.43	0.88	0.80	0.71	**0.00***	0.44	0.31	0.15	0.50
1993:12	0.14	0.39*	0.40*	0.92	0.88	0.88	**0.00***	0.61	0.55	0.23	0.50
1991:01	0.24	0.72	0.48	0.97	0.88	0.88	**0.00***	0.72	0.40	0.10*	0.54
1985:01	0.14	0.65	0.59	0.74*	0.65*	0.71	**0.00***	0.24*	0.26	0.40	0.44
1975:01	0.13*	0.45	0.52	0.76	0.67	0.67*	**0.00***	0.32	0.13*	0.19	0.38*

2007:01 - 2011:12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	-0.13*	0.43	0.13	0.82	0.79	0.74	**0.02***	-0.32	0.76	-0.10*	0.31
121-RW	-0.25	0.24	0.22	0.23	0.24	0.24	**0.02***	-0.19	0.16	-0.63	0.06
1993:12	-0.30	-0.19	-0.02	0.08	0.16	0.16	0.02*	-0.06*	**-0.01***	-0.85	-0.10
1991:01	-0.30	0.34	**0.00***	0.24	0.23	0.23	0.02*	0.10	-0.01*	-0.95	-0.01*
1985:01	-0.52	0.31	-0.07	-0.31	-0.23	-0.22	**0.02***	-0.52	-0.25	-0.34	-0.21
1975:01	-0.39	0.13*	-0.04	-0.04*	-0.05*	-0.05*	**0.02***	-0.23	-0.21	-1.00	-0.19

2012m01 - 2016m12											
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.57	1.26	0.73	1.61	1.40	1.30	**-0.02***	0.97	0.59	1.33	0.97
121-RW	0.56*	1.26	0.64*	1.53*	1.37*	1.18*	**-0.02***	1.07	0.45*	0.93*	0.94*
1993:12	0.58	0.97	0.83	1.75	1.59	1.59	**-0.02***	1.29	1.10	1.31	1.10
1991:01	0.77	1.10	0.95	1.69	1.54	1.54	**-0.02***	1.34	0.82	1.15	1.09
1985:01	0.81	0.98	1.24	1.79	1.54	1.64	**-0.02***	1.01	0.77	1.13	1.09
1975:01	0.65	0.77*	1.08	1.56	1.38	1.38	**-0.02***	0.88*	0.47	1.38	0.95

Notes: see notes from Table 1.

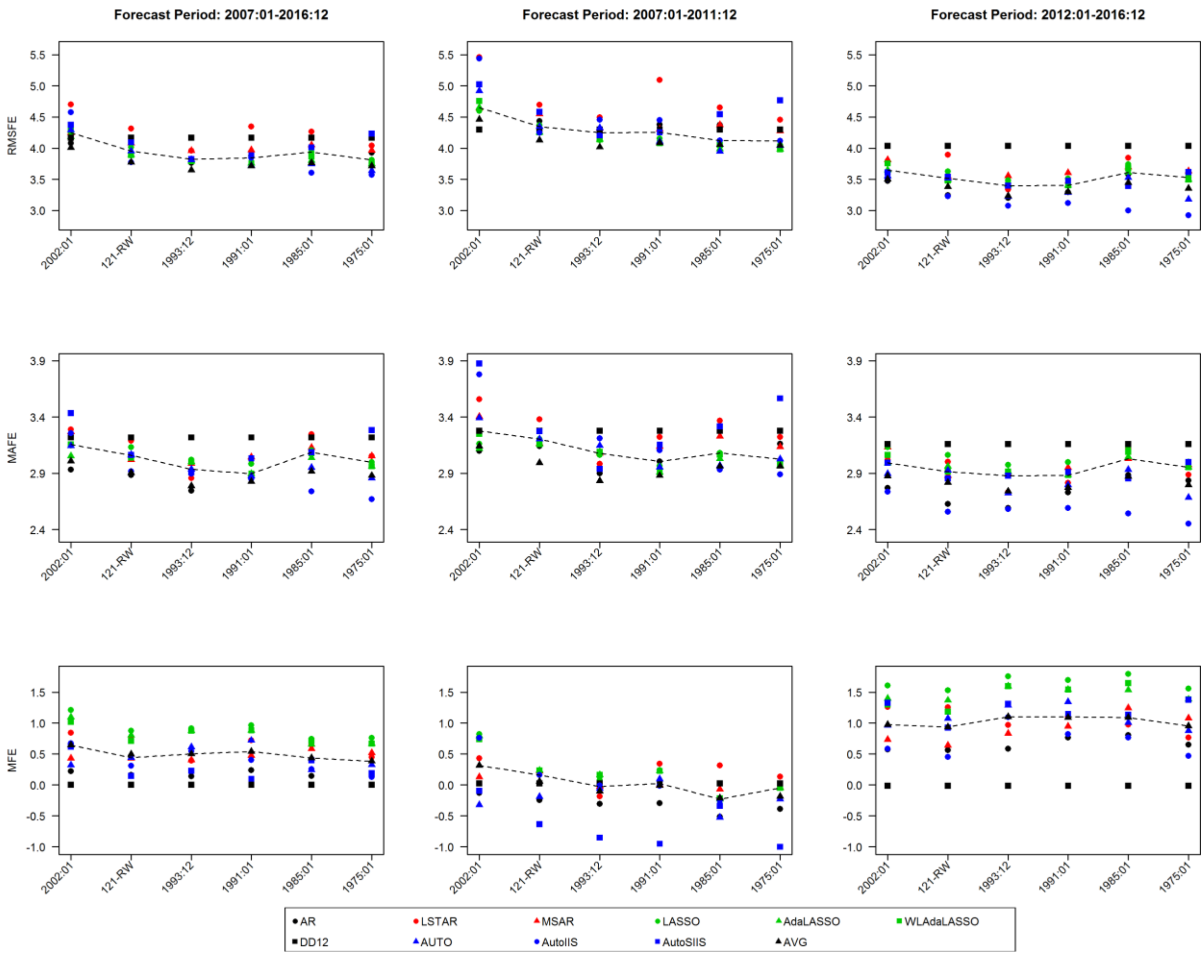


Figure 6: RMSFE, MAFE and MFE results for each estimation sample. The dashed line connects the median of these values.

than the absolute forecast error. The abrupt changes in 2007:01-2011:12 causes most models to, even at one-step ahead forecasts, miss the actual value by a relatively large amount, specially in turning points.

The MFE results in the forecast horizon's second half, however, show an upward bias, except for DD12 (which is consistent with theory, as seen in section 3.4), presenting a near zero forecast bias, whereas, in the first half, models present upward and downward biases, with median MFE nearer zero. Although forecasts are less precise in 2007:01-2012:12, the PIM-PF annual difference plunge and soar in this period causes them to deviate up and down from actual values in similar magnitude. Note that, as expected, forecasts from LASSO and derived methods tend to have larger bias than from the other models.

Consider now how forecast accuracy changes as we enlarge the initial estimation sample. The median of RMSFE and MAFE values, for both halves of the forecast horizon, tends to become lower if the sample is extended at least to 1993:12, and higher (or at least in the same level) if it is extended further, although forecasts using the sample starting in 1985:01 tend to be less precise than using that starting in 1975:01. In 2012:01-2016:12, variability between

the RMSFE (and MAFE) of different models increases for the larger estimation samples. In 2007:01-2011:12, variability of accuracy from the different models is the greatest using the sample starting in January 2002.

These differences, as we extend the estimation sample, in the median forecast performance of the models are, however, small. The median RMSFE for the whole forecast horizon, for example, changes from 4.22, using the sample starting in January 2002, to 3.82, using the expanding sample (*i.e.*, not using rolling windows) starting in December 1993.

Considering the relative performance of the models, reported in Tables 1 and 2, for the whole forecast horizon, the most accurate model depends on which estimation sample is used. Autometrics with Impulse Indicator Saturation (AutoIIS) using the sample starting in 1975:01 gives the best forecast accuracy, for both RMSFE and MAFE. For smaller samples, however, its accuracy tends to deteriorate, being inferior to either the average of forecasts from all models (AVG) or the benchmark AR.

For the first half of the forecast horizon, a period with abrupt shifts in the series, AVG delivers the best results for the 121 months rolling window estimation scheme, and for the samples starting in 1993:12 and 1991:01. For the remaining samples, the result depends on what loss function we analyze. DD12 returns the lowest RMSFE and AR the lowest MAFE using the sample starting in 2002:01. Using the sample starting in 1985:01, the minimum RMSFE is attained by Autometrics without any dummy saturation (AUTO), and, starting in 1975:01, by both AdaLASSO and WLAdaLASSO models. AutoIIS returns the lowest MAFE for both of these samples.

For the second half of the forecast horizon, a period with relatively less volatility than the first half, AutoIIS performs better than all other models, except when using the smallest sample, in which the benchmark AR gives a slightly lower RMSFE.

5.1 Diebold and Mariano Tests

Up to this point, we reported the Root Mean Squared Forecast Error and Mean Absolute Forecast Error to measure predictive performance for different models and estimation samples. However, to assess statistical differences in forecast accuracy, we perform, as described in 4.3, the Diebold and Mariano test of equal predictive performance. The analysis is separated in two subsections. The first tests, for each model, if forecasts using different samples are more accurate than the those using the latest series available, *i.e.*, that starting in January 2002. The second tests if, for each sample, any particular model have better predictive performance than the benchmark Autoregressive model of order p. Augmented Dickey-Fuller tests of the loss function differentials used in each performed DM tests are presented in the Appendix.

5.1.1 Comparing Forecasts Using Different Samples

Tables 4 and 5 report Diebold and Mariano (DM) tests with quadratic and absolute loss functions. The null hypothesis is that of equal predictive accuracy between forecasts using a particular sample and the sample starting in January 2002. The one-sided alternative hypothesis is that forecasts using the sample starting in 2002 is less accurate than those using

the other sample. We do not report the results for DD12, since its forecasts are invariant to the sample size.

DM test results clearly shows that the predictions from AR and LSTAR are improved if we use the expanding sample starting in 1993:12 instead of 2002:01, a result that is robust to the loss function used and fraction of the forecast horizon analyzed. The improvement from using the remaining estimation samples, however, is less clear cut — the null is not rejected in the majority of the cases, *i.e.*, depending on which loss function and forecast horizon is being considered.

Using larger samples with the Markov Switching Autoregressive model (MSAR) also tends to produce more accurate forecasts, particularly considering the second half of the forecast horizon. For the first half, however, the DM test using the absolute loss function rejects the null for the samples starting in 1991:01 and 1993:12, but fails to reject the null for all samples using the quadratic loss function.

Results for Lasso related methods are also not robust to the loss function considered. Although the DM test null is rejected for a number of different settings using the quadratic loss function, only WLAdaLASSO predictions with estimation sample beginning in January 1991 is statistically more accurate than those using the latest PIM-PF series at a 5% level of significance with an absolute value loss function of forecast errors.

Considering models applying the Autometrics algorithm, AutoIIS, in particular, shows robust results pointing to improvements in forecast performance from using larger samples — estimation using larger samples tends to more strongly reject the null (note, from Tables 1 and 2, that RMSFE and MAFE tend to decrease with the sample size). Autometrics without any dummy saturation and with both step and impulse indicator saturation (AUTO and AutoSIIS, respectively) produce more erratic results, although one can clearly benefit from using larger samples in a number of settings.

Using the sample starting in 1991:01 and 1993:12 also tends to improve results for the average of forecasts (AVG) in relation to using the sample starting in January 2002.

5.1.2 Comparing Forecasts Using Different Methods

Tables 6 and 7 report DM tests with quadratic and absolute loss functions, respectively. The null hypothesis is that of equal predictive accuracy between forecasts from each method and the benchmark AR model. The one-sided alternative hypothesis is that AR forecasts are less accurate than those produced by other method.

In general, when using the 121 month rolling window estimation scheme and estimation samples smaller than that starting in 1985:01, evaluated methods do not produce more accurate forecasts than those from AR. With a 5% level of significance, the one exception is the average of forecasts from all models considering only the first half of the forecast horizon, using the series starting in January 1991 and the quadratic loss function in the DM test.

Consider the results for estimation using the samples starting in January 1975 and 1985. For the first half of the forecast horizon, the DM test, using both loss functions, points that AVG predictions are more accurate than those from AR. For AutoIIS, the null from the test is rejected only when the quadratic loss function is used to construct the loss function differential series.

In the second half of the forecast horizon, the test points that AutoIIS forecasts are

Table 4: Diebold Mariano Test Comparing Samples: $g(e_{it}) = e^2_{it}$

	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	AUTO	AutoIIS	AutoSIIS	AVG
2007:01 - 2016:12										
121-RW	0.03**	0.19	0.15	0.19	0.10*	0.04**	0.05**	0.02**	0.06*	0.05**
1993:12	0.00**	0.05**	0.05**	0.04**	0.05**	0.02**	0.02**	0.01**	0.00**	0.01**
1991:01	0.05**	0.17	0.07*	0.05**	0.03**	0.01**	0.01**	0.01**	0.02**	0.03**
1985:01	0.31	0.15	0.12	0.19	0.12	0.09*	0.07*	0.01**	0.05**	0.12
1975:01	0.18	0.06*	0.07*	0.10*	0.08*	0.04**	0.03**	0.01**	0.28	0.09*
2007:01 - 2011:12										
121-RW	0.13	0.15	0.37	0.27	0.16	0.11	0.05**	0.03**	0.05**	0.10*
1993:12	0.03**	0.10*	0.14	0.11	0.10*	0.06*	0.06*	0.02**	0.01**	0.06*
1991:01	0.12	0.27	0.15	0.12	0.08*	0.05**	0.03**	0.03**	0.03**	0.08*
1985:01	0.21	0.13	0.18	0.19	0.13	0.11	0.07*	0.02**	0.10*	0.14
1975:01	0.12	0.08*	0.14	0.15	0.12	0.08*	0.07*	0.03**	0.27	0.14
2012:01 - 2016:12										
121-RW	0.07*	0.60	0.02**	0.02**	0.05**	0.01**	0.39	0.08*	0.37	0.07*
1993:12	0.04**	0.03**	0.03**	0.03**	0.06*	0.04**	0.07*	0.04**	0.13	0.02**
1991:01	0.12	0.08*	0.05**	0.01**	0.03**	0.02**	0.10*	0.02**	0.20	0.04**
1985:01	0.74	0.57	0.20	0.47	0.39	0.34	0.44	0.02**	0.11	0.31
1975:01	0.63	0.21	0.09*	0.07*	0.11	0.07*	0.02**	0.01**	0.51	0.06*

Notes: Diebold-Mariano test p-values. $g(e_{it})$ specifies the loss function used in the test. The first column denotes the initial observation of the estimation sample, 121-RW denotes a 121 months rolling window forecasting scheme. Alternative hypothesis is that forecasts from the same method using sample starting in 2002:01 are less accurate than using the sample given by the respective line. Significance at 10% and 5% levels are marked by * and **, respectively.

Table 5: Diebold Mariano Test Comparing Samples: $g(e_{it}) = |e_{it}|$

	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	AUTO	AutoIIS	AutoSIIS	AVG
2007:01 - 2016:12										
121-RW	0.26	0.32	0.01**	0.45	0.52	0.18	0.26	0.01**	0.02**	0.07*
1993:12	0.01**	0.01**	0.00**	0.15	0.35	0.14	0.08*	0.01**	0.00**	0.00**
1991:01	0.24	0.06*	0.01**	0.09*	0.13	0.04**	0.04**	0.01**	0.02**	0.01**
1985:01	0.93	0.42	0.10*	0.40	0.46	0.33	0.15	0.00**	0.04**	0.19
1975:01	0.75	0.11	0.02**	0.16	0.29	0.13	0.04**	0.00**	0.24	0.10*
2007:01 - 2011:12										
121-RW	0.63	0.31	0.08*	0.58	0.57	0.35	0.19	0.04**	0.02**	0.12
1993:12	0.09*	0.04**	0.02**	0.32	0.45	0.25	0.17	0.01**	0.01**	0.02**
1991:01	0.28	0.13	0.05**	0.18	0.20	0.10*	0.06**	0.01**	0.01**	0.03**
1985:01	0.88	0.29	0.15	0.39	0.39	0.28	0.11	0.01**	0.06*	0.16
1975:01	0.67	0.14	0.06*	0.27	0.31	0.17	0.12	0.01**	0.22	0.16
2012:01 - 2016:12										
121-RW	0.07*	0.47	0.02**	0.19	0.38	0.07*	0.58	0.07*	0.22	0.14
1993:12	0.03**	0.05**	0.03**	0.13	0.31	0.18	0.12	0.16	0.24	0.06*
1991:01	0.36	0.15	0.06*	0.11	0.20	0.10*	0.24	0.16	0.31	0.09*
1985:01	0.80	0.66	0.23	0.50	0.65	0.56	0.59	0.12	0.22	0.49
1975:01	0.69	0.29	0.10*	0.14	0.39	0.25	0.06*	0.03**	0.51	0.19

Notes: see notes from Table 4.

Table 6: Diebold Mariano Test Comparing Methods: $g(e_{it}) = e^2_{it}$

2007:01 - 2016:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.95	0.92	0.67	0.64	0.77	0.63	0.80	0.92	0.90	0.32
121-RW	0.97	0.97	0.79	0.52	0.50	0.85	0.64	0.26	0.84	0.19
1993:12	0.98	0.98	0.62	0.56	0.56	0.95	0.62	0.65	0.61	0.14
1991:01	0.94	0.78	0.43	0.25	0.25	0.88	0.22	0.41	0.51	0.07^{*}
1985:01	1.00	0.82	0.36	0.15	0.23	0.81	0.08^{*}	0.01^{**}	0.54	0.01^{**}
1975:01	0.97	0.83	0.21	0.13	0.13	0.90	0.04^{**}	0.01^{**}	0.91	0.01^{**}
2007:01 - 2011:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.91	0.58	0.49	0.54	0.63	0.26	0.78	0.92	0.89	0.29
121-RW	0.83	0.79	0.40	0.24	0.24	0.38	0.39	0.26	0.67	0.07^{*}
1993:12	0.93	0.71	0.34	0.36	0.36	0.55	0.62	0.79	0.44	0.08^{*}
1991:01	0.92	0.34	0.21	0.16	0.16	0.42	0.20	0.63	0.35	0.05^{**}
1985:01	0.95	0.51	0.13	0.09^{*}	0.11	0.41	0.07^{*}	0.14	0.76	0.02^{**}
1975:01	0.94	0.43	0.13	0.10^{*}	0.10^{*}	0.52	0.15	0.20	0.92	0.04^{**}
2012:01 - 2016:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.97	0.98	0.93	0.83	0.95	0.99	0.67	0.55	0.69	0.58
121-RW	0.95	0.98	0.99	0.94	0.96	1.00	0.95	0.45	0.89	0.92
1993:12	0.96	1.00	0.92	0.83	0.83	1.00	0.54	0.27	0.80	0.61
1991:01	0.88	0.98	0.92	0.75	0.75	0.99	0.49	0.12	0.83	0.52
1985:01	0.99	0.98	0.79	0.56	0.67	0.98	0.41	0.02^{**}	0.22	0.16
1975:01	0.92	0.97	0.56	0.44	0.44	0.99	0.08^{*}	0.01^{**}	0.63	0.09^{*}

Notes: Diebold-Mariano test p-values. $g(e_{it})$ specifies the loss function used in the test. The first column denotes the initial observation of the estimation sample, 121-RW denotes a 121 months rolling window forecasting scheme. Alternative hypothesis is that forecasts from the AR model are less accurate than those from the model given by the respective column using the same sample. Significance at 10% and 5% levels are marked by * and **, respectively.

Table 7: Diebold Mariano Test Comparing Methods: $g(e_{it}) = |e_{it}|$

2007:01 - 2016:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.99	1.00	0.89	0.76	0.91	0.95	0.90	0.95	0.99	0.79
121-RW	0.98	0.94	0.98	0.95	0.94	0.97	0.93	0.62	0.85	0.61
1993:12	0.98	0.99	0.98	0.96	0.96	1.00	0.92	0.87	0.82	0.70
1991:01	0.89	0.97	0.82	0.58	0.58	0.98	0.52	0.42	0.85	0.29
1985:01	0.95	0.77	0.54	0.34	0.45	0.77	0.15	0.01^{**}	0.48	0.01^{**}
1975:01	0.86	0.93	0.50	0.37	0.37	0.92	0.13	0.01^{**}	0.93	0.04^{**}
2007:01 - 2011:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.96	0.96	0.58	0.53	0.69	0.75	0.85	0.97	0.99	0.60
121-RW	0.90	0.58	0.63	0.54	0.54	0.70	0.64	0.79	0.68	0.10^{*}
1993:12	0.83	0.90	0.78	0.81	0.81	0.94	0.90	0.92	0.55	0.28
1991:01	0.84	0.83	0.42	0.31	0.31	0.86	0.40	0.72	0.71	0.15
1985:01	0.66	0.15	0.08^{*}	0.06^{*}	0.09^{*}	0.46	0.04^{**}	0.02^{**}	0.53	0.00^{**}
1975:01	0.72	0.29	0.11	0.11	0.11	0.69	0.20	0.05^{**}	0.90	0.01^{**}
2012:01 - 2016:12										
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.94	0.98	0.98	0.90	0.98	0.96	0.79	0.43	0.81	0.86
121-RW	0.96	0.96	1.00	0.99	0.99	0.98	0.97	0.29	0.85	0.97
1993:12	0.97	0.99	0.98	0.96	0.96	0.99	0.77	0.47	0.92	0.92
1991:01	0.78	0.96	0.96	0.83	0.83	0.96	0.65	0.16	0.85	0.68
1985:01	1.00	0.99	0.92	0.79	0.86	0.89	0.59	0.06^{*}	0.44	0.46
1975:01	0.91	1.00	0.85	0.76	0.76	0.94	0.23	0.03^{**}	0.77	0.36

Notes: see notes from Table 6.

statistically more accurate than AR forecasts for both loss functions, with 5% significance level for the largest sample, whereas the DM test for AVG forecasts is unable to reject the null using the quadratic loss function.

For the complete forecast horizon, 2007:01-2016:12, DM test null is rejected for AutoIIS and AVG for both loss functions[14].

5.2 Additional Comments

As a general result, all models tend to benefit, in terms of forecast accuracy, from using chained older PIM-PF series rather than the latest series within the current methodological benchmark, at least up to a point.

Forecasts performed using an expanding estimation sample beginning in 1993:12 show better results than those using a sample beginning in 1991:01. This may be due to the fact that the 1991-1992 period is extremely volatile, with values ranging from -20% to +30% in industrial production 12 month percentage variation, adding noise to the estimation of the latter sample.

Additionally, when estimation applying Autometrics is done with larger series, specially with impulse dummy saturation, starting in January 1975 and 1985, forecast accuracy is improved, whereas performance from other models generally worsens. In fact, AutoIIS performance improves almost monotonically with estimation sample size.

Our results point, however, that Autometrics with impulse dummy saturation may not be the best choice when forecasting in abnormally volatile periods (or in the presence of possible structural breaks), as the 2007:01-2011:12 forecast horizon in our study. Despite still performing relatively well, it fails to reject the Diebold and Mariano test null (with quadratic loss function) of equal predictive ability in relation to those from the AR model. Although forecasts from all models deteriorate in such cases, AdaLASSO/WLAdaLASSO, Autometrics without dummy saturation, and the simple average of all forecasts return equal or better results than Autometrics with impulse indicator saturation (in terms of RMSFE) in this paper's empirical exercise, being possible alternative candidates.

6 Concluding Remarks

In this paper, we analyzed how performance in forecasting the Brazilian Industrial Production Index is influenced by the use of samples of different lengths. The complete series, starting in January 1975, contains methodological updates and periods of high inflation, which could cause the weighting structure of the index to be inefficient. Additionally, long series may contain information that is no longer relevant in forecasting future values. On the other hand, using only the current series, starting in January 2002, may harm parameter estimation due to small sample size.

Annually differenced series were used to avoid problems of non-stationarity and seasonality. We separate the forecast horizon in two, 2007:01-2011:12 and 2012:01-2016:12. The first

[14]Test statistic p-value for AVG using the largest sample, starting in 1975:01, is actually 0.07, so the null would only be rejected at a 10% significance level.

is characterized by a great volatility in industrial production, caused by the subprime crisis. The second shows a less abrupt, yet persistent, decrease in industrial activity.

We assessed one-step ahead predictions from different models, re-estimating and re-specifying the models at each step. Besides the benchmark Autoregressive Model of order *p* and a naive method robust to structural breaks, we analyzed forecasts from nonlinear and data reduction techniques, and also from the simple average from all models. The relative performance between each model and the benchmark autoregressive model is also assessed for each sample.

Results show that Autometrics lag selection with impulse dummy saturation forecasting performance is improved almost monotonically with sample size — the minimum RMSFE and MAFE results, in the 2007:01-2016:12 forecast horizon, are achieved using such method with the largest series available, starting in 1975. The majority of remaining models benefit from expanding the estimation sample beginning at least up to 1993:12. Further enlarging the series may deteriorate forecast performance.

Forecast performance of AR model is superior to all other methods using the estimation sample beginning in January 1991 and smaller samples, and also using a 121 months rolling window (only the average of all forecasts is statistically more accurate for a 5% significance level in a Diebold Mariano test using a quadratic loss function). For estimation starting in January 1975 and 1985, predictions from Autometrics with impulse dummy saturation and the average of forecasts are statistically more accurate than those from AR. However, the average of predictions performs better in the first half of the forecast horizon and Autometrics performs better in the second half.

This hints that, controlling for outliers, industrial production dynamics stays relatively stable, allowing one to benefit from large samples. Using large samples with standard models, however, may impair forecast performance owing to estimation error rather than because of heterogeneity in industrial production index methodology.

References

Arango, L. and Gonzalez, A. (2001). Some evidence of smooth transition nonlinearity in Colombian inflation. *Applied Economics*, 33(2):155–162.

Bontemps, C. and Mizon, G. E. (2003). Congruence and encompassing. In Stigum, B., editor, *Econometrics and the Philosophy of Economics*, pages 354–378. Princeton University Press, New Jersey.

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Castle, J. L. and Hendry, D. F. (2014). Semi-automatic nonlinear model selection. In Haldrup, N., Meitz, M., and Saikkonen, P., editors, *Essays in Nonlinear Time Series Econometrics*, chapter 7, pages 163–197. Oxford University Press, Oxford.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

Clements, M. and Hendry, D. (1998). *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Doornik, J. A. (2009). Autometrics. In Castle, J. L. and Shephard, N., editors, *The Methodology and Practice of Econometrics*, pages 88–121. Oxford University Press, Oxford.

Engel, C. (1994). Can the Markov switching model forecast exchange rates? *Journal of International Economics*, 36(1):151–165.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Garcia, M. G., Medeiros, M. C., and Vasconcelos, G. F. (2017). Real-time inflation forecasting with high-dimensional models: The case of Brazil. *International Journal of Forecasting*, 33(3):679–693.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46(6):1303–1310.

Góes, M. C. (2005). Transition to an integrated system of business surveys — the Brazilian case. In *International Workshop on Economic Census, Beijing, China*. Available at <https://unstats.un.org/unsd/newsletter/unsd_workshops/country/Brazil_Transition%20to%20an%20integrated%20system%20of%20business%20surveys.pdf>.

Granger, C. W. and Teräsvirta, T. (1993). *Modelling Non-Linear Economic Relationships*. Oxford University Press, Oxford.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1):39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press, Oxford.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics*, 135(1):399–426.

Hendry, D. F. (2012). Mathematical models and economic forecasting: Some uses and misuses of mathematics in economics. In Dieks, D., Gonzalez, W. J., Hartmann, S., Stöltzner, M., and Weber, M., editors, *Probabilities, Laws, and Structures*, pages 319–335. Springer, New York.

Hendry, D. F. and Johansen, S. (2012). Model discovery and Trygve Haavelmo's legacy. Discussion Paper 598, University of Oxford, Oxford.

Hendry, D. F. and Krolzig, H.-M. (1999). Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez. *The Econometrics Journal*, 2(2):202–219.

Hendry, D. F. and Krolzig, H.-M. (2005). The properties of automatic GETS modelling. *The Economic Journal*, 115(502):C32–C61.

Hendry, D. F. and Mizon, G. E. (2011). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics*, 3(1). Article 1.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoover, K. D. and Perez, S. J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The Econometrics Journal*, 2(2):167–191.

IBGE (1991). *Indicadores Conjunturais da Indústria: Produção, Emprego e Salário*, volume 11 of *Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 1 edition.

IBGE (1996). *Indicadores Conjunturais da Indústria: Produção, Emprego e Salário*, vol- ume 11 of *Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2 edition.

IBGE (2003). *Classificação Nacional de Atividades Econômicas - CNAE: Versão 1.0*. Insti- tuto Brasileiro de Geografia e Estatística, Rio de Janeiro.

IBGE (2004). *Indicadores Conjunturais da Indústria: Produção*, volume 31 of *Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 1 edition.

IBGE (2015a). *Classificação Nacional de Atividades Econômicas - CNAE: Versão 2.0*. In- stituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2 edition.

IBGE (2015b). *Indicadores Conjunturais da Indústria: Produção*, volume 31 of *Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2 edition.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3):255–259.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1):1–22.

Konzen, E. and Ziegelmann, F. A. (2016). Lasso-type penalties for covariate selection and forecasting in time series. *Journal of Forecasting*, 35(7):592–612.

Lovell, M. C. (1983). Data mining. *The Review of Economics and Statistics*, 65(1):1–12.

Lundbergh, S. and Teräsvirta, T. (2005). Forecasting with smooth transition autoregressive models. In Clements, M. P. and Hendry, D. F., editors, *A Companion to Economic Forecasting*, pages 485–509. Wiley-Blackwell, New Jersey.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499.

Medeiros, M. C. and Mendes, E. F. (2016). L1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, 191(1):255–271.

Medeiros, M. C. and Vasconcelos, G. F. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138:50–52.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.

Öcal, N. and Osborn, D. R. (2000). Business cycle non-linearities in UK consumption and production. *Journal of Applied Econometrics*, 15(1):27–43.

Park, H. and Sakaori, F. (2013). Lag weighted lasso for time series model. *Computational Statistics*, 28(2):493–504.

Prince, D. d. and Marçal, E. F. (2018). Selection of lags and univariate models to forecast industrial production in an emerging country: is disagreggation useful? In *46º Encontro Nacional de Economia - ANPEC, Rio de Janeiro, Brazil*.

Santos, C., Hendry, D. F., and Johansen, S. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, 23(2):317–335.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Skalin, J. and Teräsvirta, T. (2002). Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics*, 6(02):202–241.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89(425):208–218.

Teräsvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cy- cles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7(S1):S119–S136.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Wang, H., Li, G., and Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):63–78.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In Engle, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume 4, pages 2639–2738. Elsevier, Amsterdam.

Zhang, Y., Li, R., and Tsai, C.-L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, 105(489):312–323.

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, 7(Nov):2541–2563.

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

Zou, H., Hastie, T., Tibshirani, R., et al. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5):2173–2192.

Appendix

Unit Root Tests for Loss Function Differentials

Tables A1-A4 show the p-value of Augmented Dickey-Fuller tests for unit roots for the loss function differential series used in the Diebold and Mariano tests presented in Tables 4, 5, 6 and 7 in Section 5, comparing sample and methods with quadratic and absolute loss functions.

Table A1: Augmented Dickey-Fuller Test Comparing Samples: $g(e_{it}) = e_{it}^2$.

	2007:01-2016:12									
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	AUTO	AutoIIS	AutoSIIS	AVG
121-RW	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
1993:12	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
1991:01	0.01	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.01
1985:01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
1975:01	0.01	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Notes: ADF tests include constant and trend. $g(e_{it})$ specifies the loss function used in the DM test. The first column denotes the initial observation of the estimation sample, 121-RW denotes a 121 months rolling window forecasting scheme. Minimum p-value reported is 0.01.

Table A2: Augmented Dickey-Fuller Test Comparing Samples: $g(e_{it}) = |e_{it}|$.

	2007:01-2016:12									
	AR	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	AUTO	AutoIIS	AutoSIIS	AVG
121-RW	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1993:12	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1991:01	0.01	0.01	0.01	0.01	0.01	0.01	0.10	0.01	0.01	0.01
1985:01	0.03	0.01	0.04	0.01	0.01	0.01	0.04	0.01	0.01	0.01
1975:01	0.02	0.01	0.01	0.01	0.01	0.01	0.04	0.01	0.01	0.01

Notes: see notes from Table A1.

Table A3: Augmented Dickey-Fuller Test Comparing Methods: $g(e_{it}) = e_{it}^2$.

	2007:01-2016:12									
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
121-RW	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1993:12	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1991:01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1985:01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1975:01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Notes: see notes from Table A1.

Table A4: Augmented Dickey-Fuller Test Comparing Methods: $g(e_{it}) = |e_{it}|$.

	2007:01-2016:12									
	LSTAR	MSAR	LASSO	AdaLASSO	WLAdaLASSO	DD12	AUTO	AutoIIS	AutoSIIS	AVG
2002:01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
121-RW	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1993:12	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1991:01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1985:01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1975:01	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Notes: see notes from Table A1.

Plots of the Forecasts

Figures A1-A6 plot the point forecasts of all models for the different estimation samples considered. The average of all forecasts is represented by the red lines.

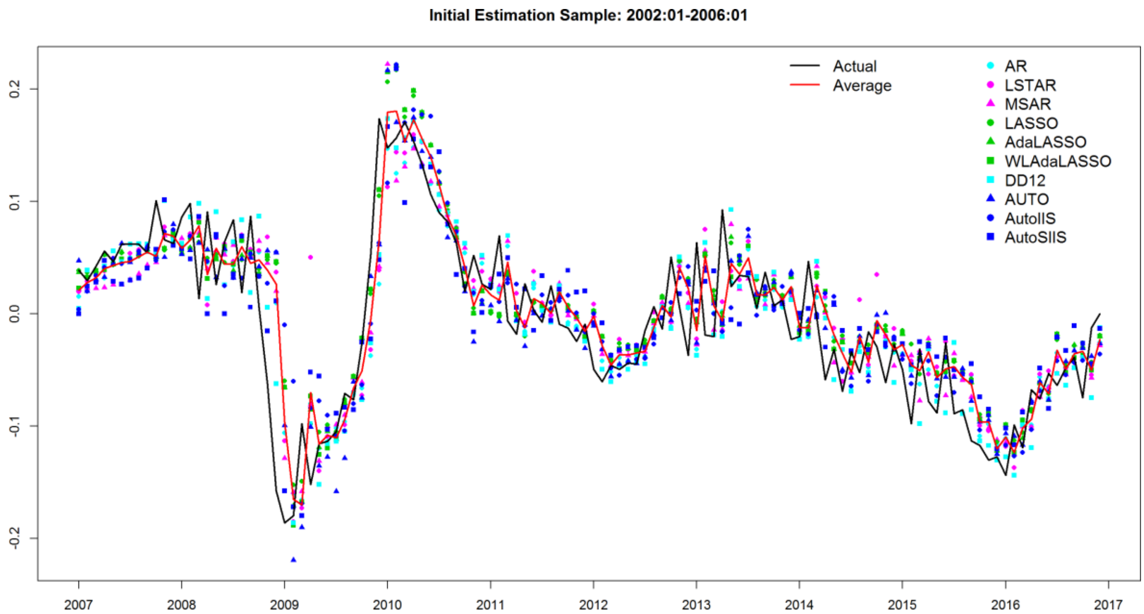


Figure A1: Point forecasts for all models for the estimation samples starting in 2002:01.

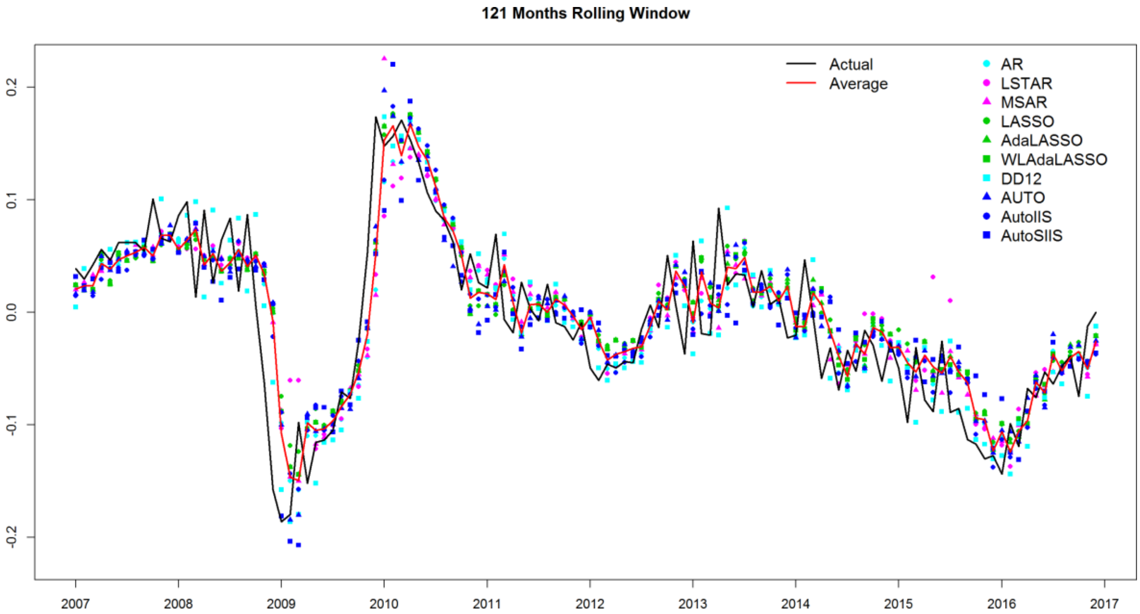


Figure A2: Point forecasts for all models for the 121 months rolling window forecasting scheme.

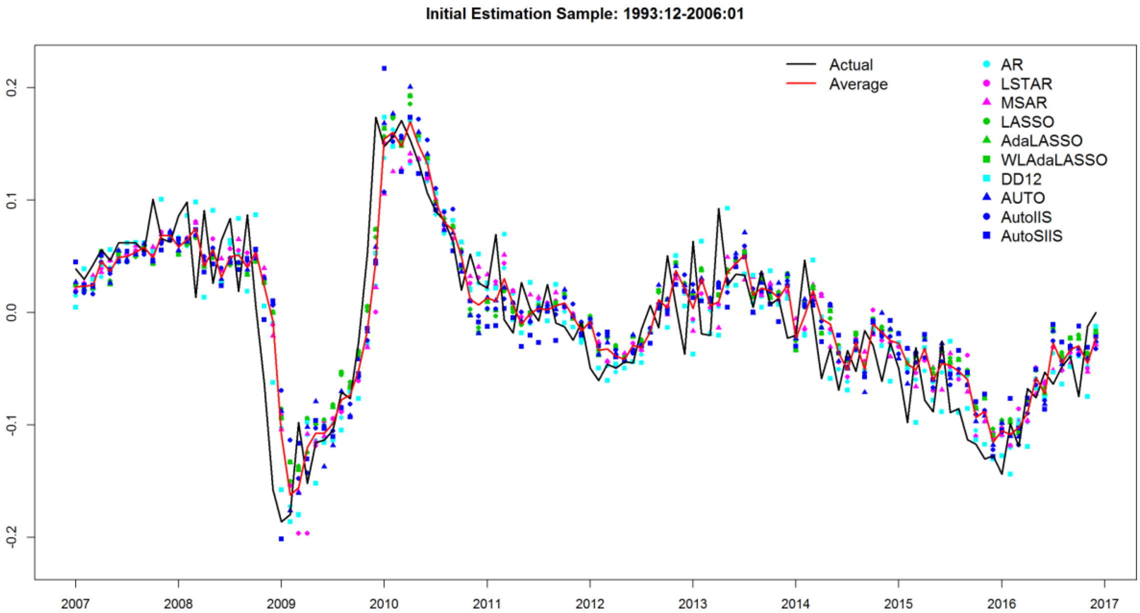


Figure A3: Point forecasts for all models for the estimation samples starting in 1993:12.

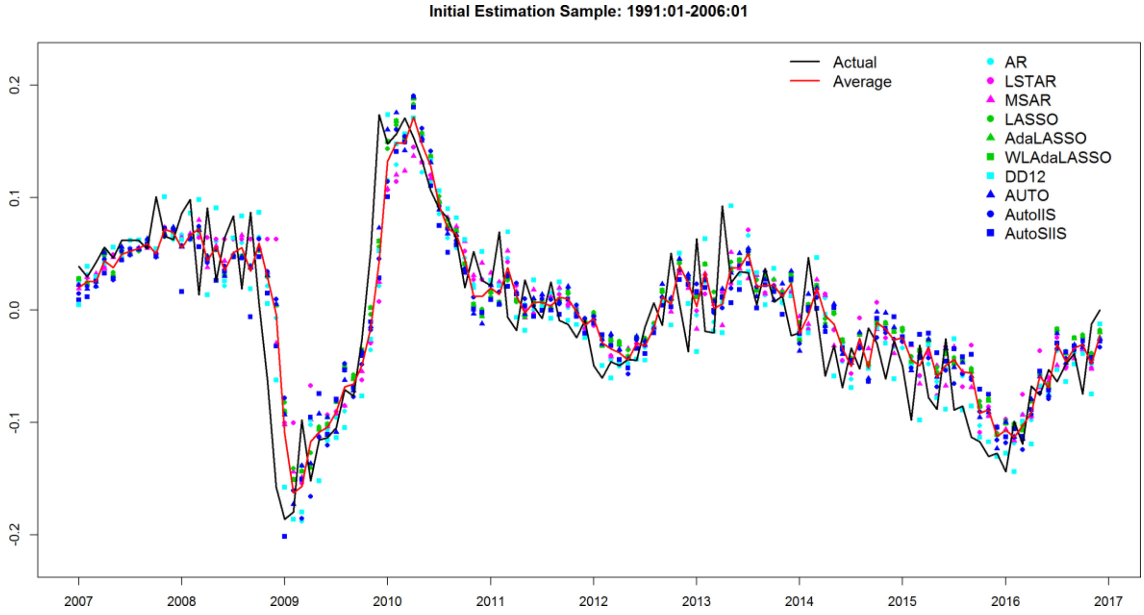


Figure A4: Point forecasts for all models for the estimation samples starting in 1991:01.

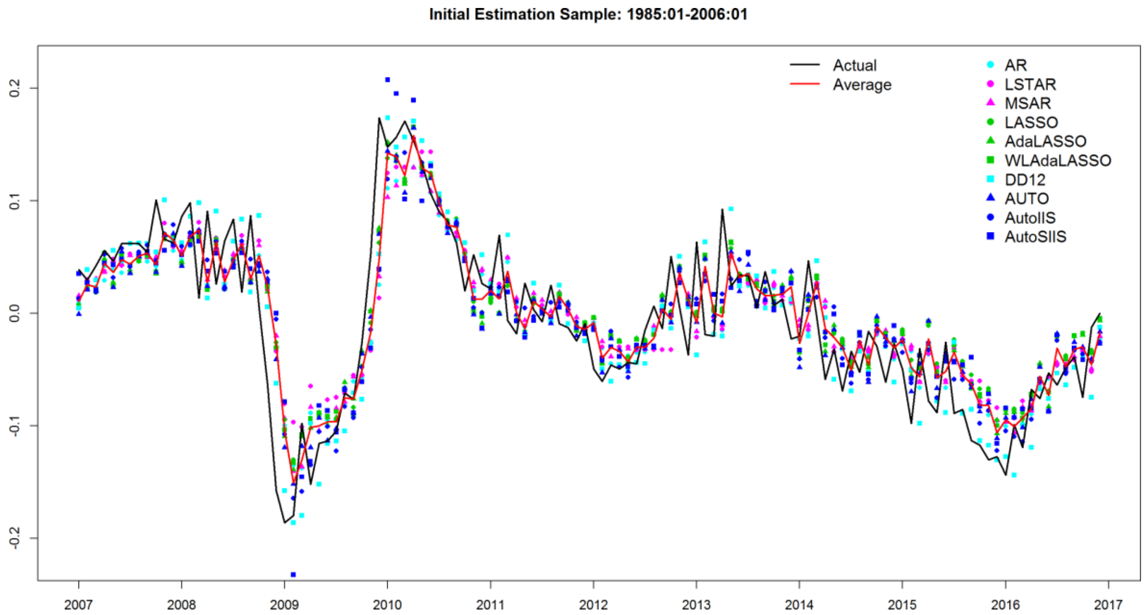


Figure A5: Point forecasts for all models for the estimation samples starting in 1985:01.

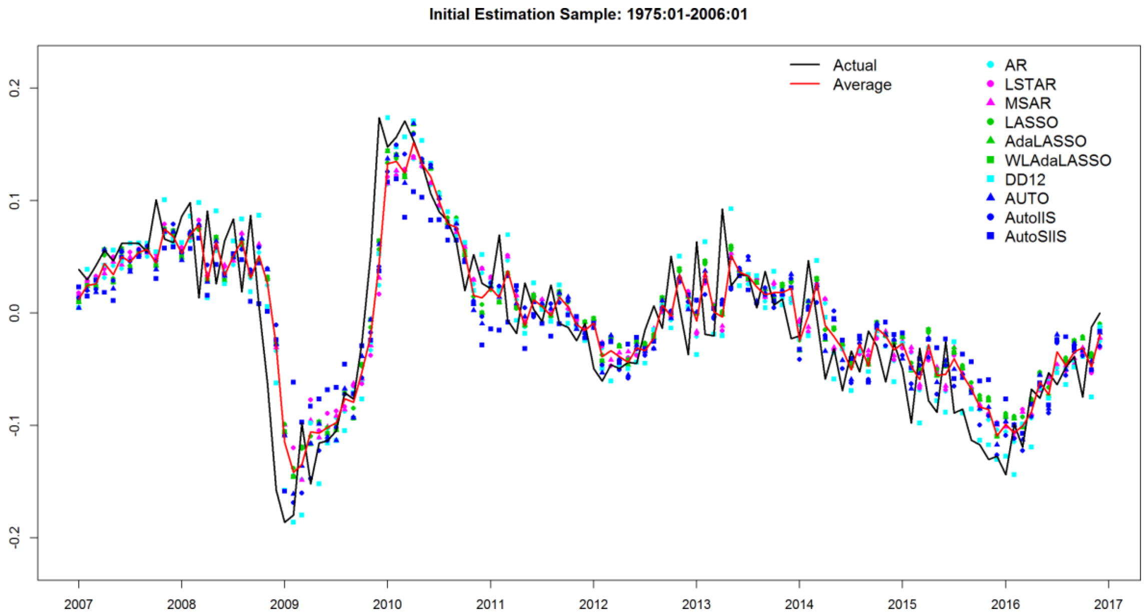


Figure A6: Point forecasts for all models for the estimation samples starting in 1975:01.

RMSFE, MAFE and MFE Results

Table A5 provides the medians plotted in Figure 6.

Table A5: Median Values of RMSFE, MAFE and MFE by Estimation Sample

	2007:01 - 2016:12			2007:01 - 2011:12			2012:01 - 2016:12		
	RMSFE	MAFE	MFE	RMSFE	MAFE	MFE	RMSFE	MAFE	MFE
2002:01	4.25	3.16	0.64	4.65	3.28	0.31	3.65	2.99	0.97
121-RW	3.95	3.06	0.44	4.34	3.20	0.16	3.51	2.91	0.94*
1993:12	3.82	2.93	0.50	4.25	3.08	-0.02*	3.40*	2.88*	1.10
1991:01	3.85	2.90*	0.54	4.25	3.01*	0.02*	3.40*	2.88*	1.10
1985:01	3.94	3.09	0.44	4.12	3.08	-0.23	3.61	3.03	1.09
1975:01	3.81*	3.00	0.38*	4.12*	3.03	-0.05	3.53	2.95	0.95

Notes: Median values for RMSFE, MAFE and MFE in Tables 1, 2 and 3. The first column denotes the initial observation of the estimation sample, 121-RW denotes a 121 months rolling window forecasting scheme. The first row marks the forecast horizon evaluated. Lowest absolute values in each column are marked by *.

4.5 Paper 5: Forecasting large realized covariance matrices: Comparing Autometrics and LASSOVAR.

Forecasting Large Covariance Matrices: Comparing Autometrics and LASSOVAR

Ronan Cunha*
Anders Bredahl Kock†
Pedro L. Valls Pereira‡

Abstract

This study aims to compare the performance of two well known automatic model selection algorithms, Autometrics (Hendry and Krolzig, 1999; Doornik, 2009), LASSOVAR and adaptive LASSOVAR (Callot et al., 2017) for modelling and forecasting monthly covariance matrices. To do so, we compose a database with daily information for 30 Brazilian stocks, which yields 465 unique entries, from July/2009 to December/2017. We apply three forecasting error measures, the model confidence set (Hansen et al., 2011) and Giacomini and White (2006) conditional test in the comparison. We also calculate the economic value for each of the forecasting strategy through a portfolio selec- tion exercise. The results show that the individual models are not able to beat the benchmark, the random walk, but a weighted combination of them is able to increase precision up to 13%. The portfolio selection exercises find that there are economic gains for using automatic model selection techniques to model and forecast the covariance matrices. Specifically, under short-selling constraint, Autometrics VAR(1) with dummy saturation delivers the highest Sharpe-ratio and economic value. When the investor is able to short-sell, ei- ther Autometrics VAR(1) with dummy saturation or adaptive Lasso VAR(1) is preferable. This final choice depend on the risk aversion of the investor. If he is less risk-averse, he prefers the former, while the latter becomes his choice if his risk-aversion sensitivity increases.

Key words: Forecasting, Covariance Matrix, Autometrics, Lasso, Vector Autoregression, Portfolio Allocation, Economic Value.

JEL Code: C32, C53, C58, G11.

*Sao Paulo School of Economics - FGV and CEQEF - FGV. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001". E-mail: cunha.ronan@gmail.com

†Oxford University, Aarhus University and CREATES, Department of Economics, Manor Road, Oxford, OX1 3UQ, UK. E-mail: anders.kock@economics.ox.ac.uk.

‡ Sao Paulo School of Economics - FGV and CEQEF - FGV. E-mail: pedro.valls@fgv.br. The third author acknowledges financial support from CNPq (309158/2016-8) and FAPESP (2013/22930-0.).

1 Introduction

Forecasting and modelling covariance matrices have attracted the interest of researchers and financist due to its great importance in portfolio selection, risk management and hedging strategies (Engle, 2009; Hlouskova et al., 2009; Boudt et al., 2013, for instance). Indeed, the classical mean-variance approach of Markowitz (1952) relies on estimating all the entries of this matrix which may become a big challenge, especially when the number of assets grows. Standard methods, such as, Baba-Engle-Kraft-Kroner (Baba et al., 1990; Engle and Kroner, 1995, BEKK) and dynamic conditional correlation (Engle, 2002, DCC) fail to deliver reliable estimates due to the curse of dimensionality (Callot et al., 2017).

Dimensionality is a serious challenge for the traditional forecasting models in this literature. The number of distinct entries grows exponentially with the number of assets. In a setup with n assets, there are $n(n+1)/2$ distinct entries in the covariance matrix. Furthermore, if we model them as a vector autoregressive of order p, VAR(p), there are $\frac{n(n+1)}{2}\left(\frac{n(n+1)}{2}p+1\right)$ parameters to estimate, including the intercepts. For instance, it means that for only 30 assets in a VAR(5), which is used in this paper, there are 465 equations and 1,081,590 parameters.

Besides the advances in estimation methodologies, such as, random matrix theory, new methods are also developed to easily construct dynamic models to the covariance matrix based on shrinkage procedures (Callot et al., 2017; Brito et al., 2018).

In this paper, we follow Callot et al. (2017) and we use the advances in model selection methodologies to deal with the curse of dimensionality aiming to accurately model and forecast all entries of the covariance matrix. Instead of using the Lasso approach as the authors, we apply the general to specific approach, Autometrics. The latter has useful improvements over other algorithms. For example, Autometrics takes the behavior of the residuals into consideration in the selection and it can easily identify structural breaks and outliers through impulse indicator saturation and its extensions (Ericsson, 2012). Outliers can seriously damage the estimation of the covariance matrix (Trucíos, Hotta, and Pereira, Trucíos et al.; Hotta and Trucíos, 2018) and its economic application (Trucíos et al., 2018).

Therefore, the main objective of this paper is to compare the forecasting performance of Autometrics and LASSOVAR when handling high dimensional system of equations. To do so, we use three different error forecasting measures as in Callot et al. (2017) and also the model confidence set (Hansen et al., 2011) and Giacomini and White (2006)'s conditional predictive ability test. We analyze the usefulness of those methods when forming investment portfolios under different environments as in Brito et al. (2018) and Callot et al. (2017). We also discuss the drivers of the dynamic of the covariance matrices over time and how they change between those methods.

The database composes daily information for 30 Brazilian stocks traded in Bovespa stock market (B3) from 2009 till 2017. Due to the promising performance found in Callot et al. (2017), we also model the logarithm of the covariance matrix (Chiu et al., 1996). Due to the lack of high frequency data for longer periods, daily data is applied to estimate the monthly covariance matrix.

The results show that it is not easy for individual models to beat the benchmark which is a random walk, but a weighted combination of them is able to increase precision up to 13%. The model confidence set show that those two methods are in all final sets and, most of the time, ranked either as first or second depending on the forecasting horizon and error measure. Giacomini and White (2006) conditional test confirms that both methods delivers equal forecasts to all forecasting horizons and error measures.

The portfolio allocation exercises show that there are economic gains replacing the no-change forecast to one of the alternative forecasting strategies, specially when short-selling is a constraint. Out of all methods, Autometrics VAR(1) with dummy saturation returns the best investment performance without short-selling measured as accumulated return. This strategy also delivers very competitive Sharpe-ratio. When shortselling is permitted, the preferable strategy depends on the investor's risk-aversion. For less risk-averse investors, Autometrics VAR(1) with dummy satu- ration has higher economic value. For more risk-averse investors, adaptive LASSO- VAR(1) returns the best value though not the highest accumulated return. However, for more highly risk-averse investors, adaptive LASSOVAR(1) is preferable if short- selling is not a restriction.

The paper is organized as following. Next section, we present the literature review and describe the two automatic selection methods compared in this study. The third section, we present the methodology and comparison strategy. In the first part of the forth section, we describe the selected models and their drivers. In the second part of this section, we present the forecasting error measures and the model confidence set results. The fifth section, there the portfolio management exercise and, the last section concludes this study.

2 Literature review

2.1 Forecasting Large Realized Covariance Matrix

Estimating, modelling and forecasting covariance matrices have been heavily studied in economics due to its substantial importance for risk management and asset allocation. In practice, the true covariance matrix is not directly observable. The traditional strategies to estimate it is to apply multivariate GARCH (Bollerslev et al., 1988) and its extensions, e.g. BEKK (Baba et al., 1990; Engle and Kroner, 1995), CCC (Bollerslev, 1990) and DCC (Engle, 2002) models. However, these mod- els impose strong restrictions on the parameters to be estimated and they heavily depends on the underlying process of the covariance matrix (Bucci, 2018). Further- more, another caveat is that they also suffer from the curse of dimensionality failing to deliver reliable estimates if the number of assets in consideration is large[1] (Callot et al., 2017).

The evolution of financial markets, has not only increased the number of assets available and also has worsened the curse of dimensionality problem, but also put available intra-day data. Merton (1980) had already showed that volatility can be

[1]There is an effort to tackle this problem in DCC in Engle et al. (2017).

defined as the sum of the squared returns at high frequency level. Andersen et al. (2001) pointed out that an observable measure of daily volatility can be obtained by summing up the squared intra-daily returns. Later, Andersen et al. (2003) showed that using the same approach it is possible to obtain all daily co-volatilities between assets, named realized covariance matrix.

The use of realized covariance has been very beneficial for investment decisions. (Fleming et al., 2003) calculated that a risk-averse investor would be willing to pay 50 to 200 basis point per year to switch from daily to intra-daily returns in order to estimate the covariance matrix.

Nevertheless, the curse of dimensionality has still been a concern. As pointed out by Bai et al. (2011) sample covariance matrix may have some undesirable properties when its dimension is large. First, when the number of asset is larger than the number of observation, the matrix is not full rank, thus it is not invertible. Second, even when it is, the expected value of its inverse is a biased estimator to its theoretical inverse. Third, the sample covariance may be very volatile and it may in- duce high weights turnover in the portfolio and it tends exceed the targeted risk out of sample. Some papers have dealt with this issue, for instance, Barndorff-Nielsen and Shephard (2004), Bickel et al. (2008), Hautsch et al. (2012), Fan et al. (2013), Lunde et al. (2016) and, for a literature review about large covariance estimation methodologies, Bai et al. (2011) and Fan et al. (2016).

When the objective turns from the estimation of the integrated covariance matrix to modelling its dynamics in order mainly to increase forecasting power, the curse of dimensionality is an issue to overcome. As highlighted previously, traditional models also suffer from the curse of dimensionality. In a set up with n assets, there are $n(n+1)/2$ distinct entries in the covariance matrix. Furthermore, if we model them as a vector autoregressive of order p, VAR(p), there are $\left(\frac{n(n+1)}{2}\right)\left(\frac{n(n+1)}{2}p+1\right)$ parameters to estimate, including the intercepts. For instance, if an investor considers a set of 30 assets to allocate in its portfolio and he models each distinct entry of the covariance matrix as a VAR(1), it means a sys- tem of 465 equations with 495 parameters to be estimated. If, instead, he uses a VAR(10), there are 2,162,715 parameters.

There are some recent papers in the realized covariance matrix forecasting literature dealing with the dimensionality problem. Bauer and Vorkink (2011) adapted the heterogeneous autoregressive (HAR) model (Corsi, 2009) to multivariate con- text. In order to reduce the number of parameters to be estimated, they apply principal components. They showed that a few latent factors are able to explain the dynamics of the covariance matrix. They calculated the co-volatilities using Barndorff-Nielsen and Shephard (2004)'s estimator and use logarithmic covariance (log Cov) of Chiu et al. (1996) to ensure positive definiteness. Chiriac and Voev (2011) developed a vector autorregressive fractionally integrated moving average (VARFIMA) model. As opposite to Bauer and Vorkink (2011), the apply Cholesky decomposition to ensure positive definiteness matrix. Their procedures if based on three steps: i) the covariance matrix is decomposed into Cholesky factors; ii) these factors are modelled and predicted using VARFIMA model and; iii) the covariance matrix is recomposed with the forecast factors. Gouriéroux et al. (2009) and its extension by Golosnoy et al. (2012) proposed the use autorregressive Wishart model

to capture the dynamic of covariance matrix. The common drawback of all those studies is they still suffer from the curse of dimensionality when the number of assets grows (Callot et al., 2017). Indeed, Golosnoy et al. (2012) pointed out that their model may be applicable to about ten assets. They also used a few assets from North America' stock market[2] in their empirical exercise.

In this contexts, there are a few recent sutdies focusing on more flexible approaches to tackle the high dimensionality problem. Callot et al. (2017) modelled the dynamic of the $n \times n$ realized covariance matrix (Σ_t) as a vector autoregression of order p. In the paper p is equal to 1, 5 and 20. Each distinct entry of the matrix is an equation which is composed of the p lags of all the other entries. Thus, the whole system has $k = n(n + 1)/2$ equations and $k\,(kp + 1)$ parameters to estimate. The authors constructed the upper bound of the covariance matrix forecasting error and how it translates to the portfolio variance forecasting error for Lasso and adaptive Lasso. They also apply the methodology to 30 assets from USA for daily, weekly and monthly aggregations. To ensure positive definiteness, they apply two strategies: i) logarithmic covariance matrix (log Cov) as in Chiu et al. (1996) and, ii) eigenvalue cleaning as in Hautsch et al. (2012). Their results show that it is not easy to beat the benchmark of no change but the log Cov outperformed all the other models more frequently in all aggregation levels. They also estimated that a risk-averse investor is willing to pay 11% per year to switch from the no change forecast to the log Cov model.

Engle et al. (2017) combined the composite likelihood method of Pakel et al. (2017) which allows the estimation of large dimensional DCC models with non-linear shrinkage method of Ledoit and Wolf (2017) derived from Random Matrix Theory. Using CRSP data, they constructed portfolios with up to 1000 assets and they showed their method outperforms alternative strategies, such as, traditional DCC models and Riskmetrics 2006 (Zumbach, 2007).

Another study that also dealt with a large number of assets, 430 stock from composing S&P500, is Brito et al. (2018). They put forward a methodology which combines asset pricing factor models, heterogeneous autoregressive models and Lasso. They argued that, although the previous procedures are able to deal with large number of assets, they become unfeasible if the number of assets keeps growing. Computationally speaking, the previous methods indeed have this requirement.

The recent studies rely on the shrinkage methods, especially adaptive Lasso (Tibshirani, 1996; Zou, 2006). However, there are alternative model selection methods which, in essence, could also be applied to model the covariance matrix similarly to Callot et al. (2017). In the next section, we present Autometrics which is an alternative method to Lasso and extensions.

2.2 Automatic Model Selection Methods

Automatic Model Selection methods have gained importance in econometrics due to the increasing amount of available data. Researchers need to deal with highly dimensional models that sometimes make the estimation totally infeasible. The

[2] Bauer and Vorkink (2011) and Golosnoy et al. (2012) used five and Chiriac and Voev (2011) used six assets from United States.. Gouriéroux et al. (2009) used three from Canada.

traditional selection methods, such as R^2 and information criteria are impracticable if the number of candidate variables is large. The space of potential final models exponentially increases with the number of candidates. If there are k candidate variables, there are 2^k possible final models.

In this paper, we compare the forecasting performance of two automatic model selection techniques. The first is Lasso and its extension, adaptive LASSOVAR, that we present in the next section and the second is Autometrics which belongs to a different class of section models called General to specific (Gets) approach developed by Hendry and Krolzig (1999); Doornik (2009).

2.2.1 Lasso for Vector Autoregressive model (LASSOVAR)

The least absolute shrinkage and selection operator, or as it is better known Lasso, was proposed by Tibshirani (1996). The methods is based on the maximization of the square residuals subject to the sum of the absolute regression coefficients which allows the selection and estimation of the parameters simultaneously. It has gained popularity due to its ability to shrink a coeffients to zero and return a sparse model (Zhao and Yu, 2006).

Besides these desired characteristics, Zou (2006) showed that Lasso is consistent under nontrivial conditions and it lacks the oracle property. Thus, Zou (2006) put forward the adaptive Lasso which fixs those problems. This method use adaptive weights to penalize different coefficients.

Lasso and adaptive Lasso have been developed under the context of a single equation to model the conditional mean. Callot et al. (2017) extended the methodology to system of equations to model all entries of the covariance matrix. They developed this method as an alternative and a more flexible approach to tackle the curse of dimensionality when dealing with an increasing number of assets.

Callot et al. (2017) modelled the dynamic of the $n \times n$ realized covariance matrix (Σ_t) as a vector autoregression of order p. Each distinct entry of the matrix is an equation which is composed of the p lags of all the other entries. Thus, the whole system has $k = n(n + 1)/2$ equations and $k\,(kp + 1)$ parameters to estimate. We can to write the system as

$$y_t = w + \sum_{i=1}^{p} \Phi_i y_{t-i} + e_t,\ t = 1, \dots, T \tag{1}$$

where $y_t = \text{vech}(\Sigma_t)$ and vech(.)[3] is the half-vectorization operator return a vector of length $n(n+1)/2$ with the unique entries of Σ_t. Φ_i is the $k \times k$ matrix of parameters of the i-th lag and w is the $n(n + 1)/2 \times 1$ vector with the intercepts.

To circumvent the dimensional problem, Callot et al. (2017) show that it is possible to apply Lasso by Tibshirani (1996) and adaptive Lasso (Zou, 2006) equation by equation to select the number of relevant regressors. Thus, $\Phi = (\Phi_1, \dots, \Phi_p)$ will be a sparse matrix.

[3]Define $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, then $\text{vech}(A) = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

For convenience, we shall rewrite equation (1) in a stacked form. Let $Z_t = (1, y_{t-1}, \ldots, y_{t-p})$ be the $kp+1$ vector of explanatory variables as time t and $Z = (Z_T, \ldots, Z_1)$ the $T \times (kp+1)$ matrix of covariates for each equation. Let $y_i = (y_{T,i}, \ldots, y_{1,i})$ be the $T \times 1$ vector of observations on the *ith* variable ($i = 1, \ldots, k$) and $e_i = (e_{T,i}, \ldots, e_{1,i})$. Finally, $\gamma_i = (w_i, \beta_i)$ is the $(kp+1)$ parameter vector of equation i. The stacked form of (1) is

$$y_i = Z\gamma_i + e_i, \; i = 1, \ldots, k \tag{2}$$

The main assumption for the estimation of model (2) is the sparsity of γ_i. Let $\hat{\gamma}_i$ be the estimator of γ_i which is the solution to the following minimisation problem[4]:

$$\hat{\gamma}_i = \operatorname{argmin} \left\{ \|y_i - Z\gamma_i\|^2 + 2\lambda_T \|\beta_i\|_{\ell_1} \right\}, \tag{3}$$

where $\|x\| = \sqrt{\sum_{i=1}^{m} x_i^2}$, $\|x\|_{\ell_1} = \sum_{i=1}^{m} |x_i|$ and λ_T is the weight of the penalty term and it is selected through the Bayesian Information Criterion (BIC). Thus, the BIC for equation i and penalty parameter λ_T is

$$BIC_i(\lambda_T) = T \times \log(\hat{e}'_{\lambda_T,i}\hat{e}_{\lambda_T}) + \sum_{j=1}^{kp} (\hat{\beta}^{\lambda}_{ij} \neq 0) \times \log(T), \tag{4}$$

where $\hat{e}_{\lambda_T,i}$ is the vector of residuals and $\hat{\beta}^{\lambda}_{ij}$ are the estimated β corresponding to penalty parameter λ_T.

For the adaptive Lasso, Callot et al. (2017) proposed to apply Lasso in a first stage to reduce the dimensionality and thus to use estimated parameters as the weights for adaptive Lasso, in the second stage.

Formally, let $J(\hat{\beta}_i) = \{j \in \mathbb{R}^{kp} : \hat{\beta}_{i,j} \neq 0\}$ be the indices of the coeffients considered nonzero by Lasso for equation i and $\tilde{J}_i(\hat{\gamma}_i) = \{1\} \cup J(\hat{\beta}_i) + 1$ where we can understand the set addition elementwise.

Let $\hat{\gamma}_i^{adap}$ be the estimates coefficients by adaptive Lasso, thus $\hat{\gamma}_i^{adap}$ is the solution to the following minimization problem:

$$\hat{\gamma}_i^{adap} = \underset{\gamma \in \tilde{J}(\hat{\gamma}_i)}{\operatorname{argmin}} \frac{1}{T} \|y_i - Z_{\tilde{J}(\hat{\gamma}_i)}\gamma_{i,\tilde{J}(\hat{\gamma}_i)}\|^2 + 2\lambda_T \sum_{j \in J(\hat{\beta}_i)} \frac{|\beta_{i,j}|}{|\hat{\beta}_{i,j}|}, \; i = 1, \ldots, k \tag{5}$$

This approach allows the application of adaptive Lasso in a setting where there are more candidate variables than observations. A drawback of this methodology is that it does not take the behavior of the residuals into consideration in the se- lection process. Thus, it is not guaranteed that the residuals of the final model are congruent. Autometrics, which we present in the next section, takes the behavior of the residuals into consideration by applying a set of diagnostic tests during the selection process.

[4]Just for comparison, in Tibshirani (1996), the minimization problem is $\|y_i - Z\gamma_i\|^2 + \lambda\|\beta_i\|_1$.

2.2.2 Autometrics

Autometrics (Doornik, 2009) is a multi-path general to specific model selection algorithm. It is an extension of Hoover and Perez (1999) and Hendry and Krolzig (1999) which are derived from the Hendry (1995)'s theory of reduction.

It is implemented in OxMetrics 8 and it follows the following step:

1. General Unrestricted Model (GUM): The GUM is the initial model which encompasses the whole information, that is, all candidate variables. It is equation (2). It is expected to encompass the parsimonious congruent models, otherwise, it will suffer with omitted variable bias (Castle and Hendry, 2010).
2. Dummy saturation: In order to control for structural breaks and outliers, Autometrics applies Impulse Indicator Saturation (IIS), Step Indicator Saturation (SIS), Zero sum pairwise IIS (DIIS) and extensions, see Ericsson (2012).
3. Tree search: this is the main difference from the multi-path strategy in Hendry and Krolzig (1999) algorithm. Chosen a significance level, each variable or a group of variables is deleted. To efficiently search, Autometrics maps all possible models and it structures the tree search in a way that avoids estimating the same models twice.
4. Highly correlated regressors: It is well known that highly correlated regres- sors are problematic for all model selection algorithms. For Autometrics, in particularly, if there is a relevant variable highly correlated with another regressor, it tends to exclude them due to inflation in standard errors. In order to avoid that important variables are removed because of that, Autometrics also searches for possible variables that should be re-included. It looks for models in the set of previous excluded variables.
5. Diagnostic test: The behavior of error is also assessed. In order to be considered a final candidate it must pass in the diagnostic tests pre-specified.
6. Tie-breaker: In the end of each branch of the tree search, there is a final candidate. They are merged and become the new GUM. The interaction goes on until the next GUM be equal to the previous one. If there are still more than one model that passes in the diagnostic test, the finalist may be either the union of them or selected by information criteria.

Its procedure is very intuitive as it has been designed to follow more rigorous and concisely the procedure researchers would do manually. As pointed out by (Castle et al., 2013), model selection may lead them to fall into the data-mining trap.

Autometrics can also handle system of equation as vector autoregressive models. However, in this study, we apply the selection methods equation by equation as the number of equations in our model is greater than the number of observation as it will be described in the next section.

Autometrics' theoretical consistency and properties have been demonstrated in some papers. Campos et al. (2003) mapped information criteria into the implicit significance levels approach which is the exclusion criterion used by Autometrics to study its statistical properties. Hendry and Krolzig (2005) showed how to de- bias the estimator after repeated and sequentially hypothesis test which truncates the t-Student distribution far from the origin, consequently, generating coefficients which are upward bias and standard error which are downward biased. Johansen

and Nielsen (2009) showed the robustness of impulse indicator saturation and they derived its asymptotic distribution.

Epprecht et al. (2019) also studied the performance of Lasso against Autometrics but in a different setup. While we compare those methods in system of equations for modelling covariance matrices, the authors analyzed univariate case for conditional mean. They reported, through simulation that when the number of variables is larger than the number of observations, both methods performed similarly. Using real data, Autometrics presented better results in-sample measured by R^2 while using a modified version of Diebold and Mariano (1995) test of predictive accuracy (Harvey et al. (1997)) Lasso, adaLasso and Autometrics are not distinguishable.

3 Methodology

3.1 Databases and construction of the covariance matrix

The database composes the daily closing prices of the 30 most liquidity stocks[5] over the period 2009 and 2017. We split the database from 2009 to 2014 for the first estimation sample and we forecast from January 2015 till December 2017. We apply a rolling window of 60 observation and we forecast up to 6 months ahead.

We compute the monthly covariance matrix (Σ_t) using daily stock returns. Let $r_{t_\tau} = \ln(p_{t_\tau}/p_{t_{\tau-1}})$ be the return of stock i in month t and day τ , where ln is the napierian logarithm. We calculate the covariance between the $i-th$ and $j-th$ stocks for month t, $\sigma_{ij,t}$, according to:

$$\sigma_{ij,t} = \frac{1}{t_d} \sum_{t_\tau=1}^{t_d} (r_{t_\tau i} - \bar{r}_i)(r_{t_\tau j} - \bar{r}_j), \tag{6}$$

where t_d is the number of days in month t and $\bar{r}_i = \frac{1}{t_d} \sum_{t_\tau=1}^{T_d} r_{t_\tau i}$ is the monthly average return of stock i.

In order to ensure that we forecast positive definite matrix, we model the logarithm of the variance (Chiu et al., 1996; Bauer and Vorkink, 2011). The reason for this choice is because Callot et al. (2017) found that the log covariance delivers the best performance if compared to other methods of ensuring positiveness, such as the eigenvalue cleaning (Hautsch et al., 2012).

3.2 Computing the Forecast

After selecting the variables and estimating the vector of parameters $\hat{\gamma}_i$ for equation i, we do the forecasts as usual in VAR models. The one step ahead forecasts are given by equation (7). The forecasts to $(T+h)$ step is done recursively using the previous periods' forecasts, as in equation (8).

$$\hat{y}_{i,T+1|T} = \hat{\gamma}_i Z_T. \tag{7}$$

[5]The complete list is in table A1 in the appendix.

$$\hat{y}_{i,T+h|T} = \hat{\gamma}_i \hat{Z}_{T+h-1|T}. \tag{8}$$

Non-linear models may produce forecasts clearly unreasonable (Swanson and White, 1995, 1997; Teräsvirta et al., 2003). To deal with it, Swanson and White (1995) came up with a filter which unrealistic forecasts are replaced by reasonable ones. They named this procedure, insanity filter. Furthermore, Kock and Teräsvirta (2014) showed that it may also improve forecasting accuracy of three automatic model selection: Autometrics, QuickNet and Marginal bridge estimator. Callot and Kock (2014) applied this filter to Lasso and adaptive Lasso.

Therefore, we also use this filter to trim insane forecasts as in Callot and Kock (2014) and Callot et al. (2017). If $\hat{y}_{i,T+h|T} \notin [y_T \pm 4 \times SD]$, where SD is the standard deviation in the estimation sample, we replace $\hat{y}_{i,T+h|T}$ with y_T. We substitute insanity for ingenuity. Kock and Teräsvirta (2014) noted that the particular choice of insanity filter is not important; what matters is to eliminate unrealistic forecasts. As the forecasts computed are recursively, we weed out the forecast outside the interval before forecasting the next period. The reason for this choice is that an unreasonable is most likely to generate another unreasonable forecast. Using this procedure, we reduce the number of replacements.

3.3 Forecasting combination

Many studies advocates that there are gains in precision when combining forecast from different models (Granger, 1969; Granger and Ramanathan, 1984), we also test its performance for the covariance matrix. Cavaleri and Ribeiro (2011) show the improvement for Brazilian volatility as well.

The explanation for forecasting combination is due to the data generator process instability that may averaged out as suggested by Diebold and Pauly (1987), Newbold and Granger (1974) and Hendry and Clements (2004). According to Samuels and Sekkel (2017), even the simplest combination methods such as equal weights would improve the precision of the predictions.

We follow Bates and Granger (1969) and we apply the inverse mean square forecasting error (MSFE). The weights are calculated in the following manner: First, we use six months to evaluate the forecasts. Then using a rolling window we compute the weights through:

$$w^{Comb}_{t,i} = \frac{MSFE^{-1}_{(t-6,t-1),i}}{\sum_{j=1}^{N} MSFE^{-1}_{t-6,t-1,j}} \tag{9}$$

3.4 Forecasting Models and benchmark

The benchmark used to compare the forecasts is the no-change volatility, that is, a random walk. Thus, the forecast is $\hat{y}_{i,T+h|T} = y_{i,T}$ for $h = \{1, ..., H\}$.

We also estimate the Multivariate Exponentially Weighted Moving Average (EWMA) with decaying parameter $\lambda = 0.97$. This model is also know as Riskmetrics 1996 (Morgan et al., 1996) which assumes the form of equation (10):

$$\Sigma_{,t} = 0.97\Sigma_{t-1} + 0.04 r_{i,t-1} r_{j,t-1}, \tag{10}$$

where $r_{i,t}$ is the $1 \times n$ vector of returns. A drawback of this model is that it does not capture long memory processes. Thus, we also compute Riskmetrics 2006 (Zumbach, 2007):

$$\Sigma_t = \sum_{i=1}^{m} \omega_i \Sigma_{i,t}, \tag{11}$$

$$\Sigma_{i,t} = (1-\lambda_i) e_{t-1} e_{t-1} + \lambda_i \Sigma_{i,t-1}, \tag{12}$$

$$\omega_i = \frac{1}{C}\left(1 - \frac{\ln(\tau_k)}{\ln(\tau_0)}\right), \lambda_i = \exp\left(-\frac{1}{\tau_i}\right), \tau_i = \tau_1 \rho^{i-1}, \quad i = 1, 2, \ldots, m$$

where C is a normalization constant which ensures that $\sum_{i=1}^{m} \omega_i = 1$, $\tau_0 = 1560$ is the decay parameters, $\tau_1 = 4$ is a lower cut-off and $\tau_{max} = 512$ is an upper cut-off, $\rho = \sqrt{2}$ is an additional parameter to operationalize the model. Zumbach (2007) suggested the values to each of the four parameters through forecasting exercises[6]. According to Sheppard (2013), $m = 1 + \frac{\ln\left(\frac{\tau_{max}}{\tau_1}\right)}{\ln(\rho)}$.

To summarize all the models we estimate in this study, in table 1 there is the list of model and a brief description of them.

Table 1: List of forecasting models

Name	Description
No-change	Random Walk: $\hat{\Sigma}_{t+1} = \Sigma_t$
EWMA	Exponentially Weighted Moving Average ($\lambda = 0.97$)
Riskmetrics 2006	Riskmetrics methodology 2006
LASSOVAR(1)	VAR(1) selecting by LASSO
AdaLASSOVAR(1)	VAR(1) selecting by adaptive LASSO
Autometrics VAR(1)	VAR(1) selecting by Autometrics
Autometrics VAR(1) DIIS	VAR(1) selecting by Autometrics using IIS, SIS and DIS.
LASSOVAR(5)	VAR(5) selecting by LASSO
AdaLASSOVAR(5)	VAR(5) selecting by adaptive LASSO
Autometrics VAR(5)	VAR(5) selecting by Autometrics
Autometrics VAR(5) DIIS	VAR(5) selecting by Autometrics using IIS, SIS and DIS
Combination equal	Forecast combination - equal weights
Combination weight	Forecast combination - weights calculated by equation (9)

Note: VAR(p) means vector autoregressive model of order p. IIS is impulse dummy saturation; SIS is step impulse saturation; DIS is difference impulse saturation.
Source: The authors.

3.5 Forecasting evaluation

Let $\hat{e}_{t+h}$ denote vech $\left(\hat{\Sigma}_{t+h} - \Sigma_{t+h}\right)$, where $\hat{\Sigma}_{t+h}$ is the forecast covariance matrix for time period $t+h$ using the methods we described in the previous sections.

[6] As the results for Riskmetrics 1996 and Riskmetrics 2006 are very similar, to save space, we show the forecasting error measures for the latter in table A2 in appendix.

Σ_{t+h} is the observed covariance matrix for the same period. The forecast are computed for periods h, where $h = 1, \ldots, H$. In order to evaluate the methods' performances, we compute the following indicators:

1. Average g_2-forecast error (L2): $\frac{1}{H}\sum_{h=1}^{H} \|\hat{e}_{t+h}\|$;
2. Average median absolute error (AMedAFE): $\frac{1}{H}\sum_{h=1}^{H} \text{median}(|\hat{e}_{t+h}|)$. It is more robust to outliers then than g_2;
3. Average maximal absolute forecast error (AMaxAFE): $\frac{1}{H}\sum_{h=1}^{H} \max(|\hat{e}_{t+h}|)$

We compute all the indicators to the whole forecast covariance matrix and individually for the variance and covariance.

3.5.1 Predictive ability tests

We also apply the model confidence set (MCS) of Hansen et al. (2011) and conditional predictive ability test of Giacomini and White (2006). The MCS consists of an algorithm that selects the best model under some criterion given a level of confidence. In this sense, it is analogous to a confidence interval for a parameter. The advantage of using this procedure is the ability it has to rank all the models given their predictive performance.

MCS uses an equivalence test, δ_M, to create a set of selected models. Formally, it tests the null hypothesis that $H_0 : E(d_{i,j}) = 0$ for all $i, j \in M$ where M is a set of selected models and $d_{i,j} = g(e_i) - g(e_j)$ where g_{ij} is some loss function. The alternative hypothesis is given by: $H_1 : E(d_{i,j}) = 0$ for all $i, j \in M$.

The null hypothesis can be tested using the following statistic which does not have a standard distribution but it can be easily simulated by bootstrap techniques:

$$T_{R,M} = \max_{i,j \in M} |t_{ij}| \tag{13}$$

where $t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{Var(\bar{d}_{ij})}}$ and the elimination rule is $\rho_M = \arg\max_{i \in M} \sup_{j \in M}(t_{ij})$ and $\bar{d}_{ij} = H^{-1}\sum_{t=1}^{H} d_{ij}$ for some i and j.

The algorithm works as following:

1. Set the initial set of models $M_0 \in M$;
2. It tests H_0 applying δ_M at a significance level α;
3. If H_0 is rejected, it eliminate the object using ρ_M and it goes back to step 1. The procedure ends when H_0 is not rejected. Then, the surviving objects compose the model confidence set, $M_{1-\alpha}$;
4. At the end, it uses the p-value to rank the surviving models.

Although the advantage of comparing multiple forecasts and the ability to rank the elements of confidence set, its assumptions may not hold when forecasts are based on estimated parameters Hansen et al. (2011). According to Hansen et al. (2011), some of the problems in this context can be avoid using a rolling window to estimate the parameters, but further modifications are also required. Otherwise, it can

compromise the results, especially when comparing nested models.

To overcome this problem, Giacomini and White (2006) generalized West (1996) and McCracken (2000) to incorporate the use of conditional information, which al-

lows to deal with parameter uncertainty and comparison of nested models. Formally, Giacomini and White (2006) tests the null hypothesis that $H_0 : E(g(y_t, f_{1t}(\hat{\beta}_{1t})) - g(y_t, f_{2t}(\hat{\beta}_{2t}))|I_t) = 0$, where g is loss function, f_{1t} and f_{2t} are two competing models and, $\hat{\beta}_{1t}$ and $\hat{\beta}_{2t}$ are its respective estimated parameters. The expectation is conditioned on the set of information I_t.

The test statistic has the form:

$$T_{\tau}^{h} = n \left(n^{-1} \sum_{t=m}^{T-\tau} h_t \Delta g_{m,t+\tau} \right) \tilde{\Omega}_n^{-1} \left(n^{-1} \sum_{t=m}^{T-\tau} h_t \Delta g_{m,t+\tau} \right) \sim \chi_q^2 \tag{14}$$

where τ is the forecast horizon, $\Delta g_{m,t-\tau} = g(y_t, f_{1t}(\hat{\beta}_{1t})) - g(y_t, f_{2t}(\hat{\beta}_{2t}))$, $n = T - \tau$, h_t is a $q \times 1$ I_t-measurable function. $\tilde{\Omega}$ is a consistent estimator of the variance of $h_t \Delta g_{m,t+\tau}$. When $\tau > 1$, Giacomini and White (2006) argued that $h_t \Delta g_{m,t+\tau}$ is serial correlated, thus, some robust estimator, such as Newey and West (1987) must be used.

This test also requires that the parameters are estimated using a fixed scheme, where it is estimated once and computed all the forecasts, or using a rolling window, which we use in this study. The idea behind this requirement is that this test can handle data that are mixing and heterogeneous (Giacomini and Rossi, 2010).

3.5 Portfolio allocation

The portfolio analysis is based on the framework developed by Fleming et al. (2001, 2003) as in Callot et al. (2017). We consider a risk-averse agent that needs to compose his investment portfolio with $n = 30$ assets. To do so, he uses the mean-variance analysis in every period t of time. He has the utility function given by

$$U(r_{pt}) = (1 + r_{pt}) - \frac{\gamma}{1+\gamma} (1 + r_{rt})^2,$$

where r_{pt} is the portfolio return at time t and γ is investor's risk averse coefficient. Larger γ means the investor is more risk-averse. We assume $\gamma = 1, 5, 10$.

Every period $t = t_0, \dots, T$, the investor decides the vector of weights for the next period $\hat{w}_{t+1}$ using information up to t solving the following problem:

$$\begin{aligned} \hat{w}_{t+1} = \underset{w_{t+1}}{\text{argmin}}\ & w_{t+1} \hat{\Sigma}_{t+1} w_{t+1} \\ \text{s.t.}\quad & w_{t+1} \hat{\mu}_{t+1} = \mu \\ & \sum_{i=1}^{n} w_{i,t+1} = 1 \\ & \sum_{i=1}^{n} |w_{i,t+1}| \mathbb{1}(w_{it} < 0) \leq 0.30 \\ & |w_{i,t+1}| \leq 0.20, \end{aligned} \tag{15}$$

where w_{t+1} is $n \times 1$ vector of portfolio weights on the stock, μ is the target return which we consider 10% per year, $\hat{\mu}$ is a $n \times 1$ vector of forecast returns computed

by a moving average of the same size of the estimation window for the covariance matrix and $\mathbb{1}(\,)$ is a indicator function.

The investor's problem is to choose the weights for next period's investment subject to the target return μ and the forecast covariance matrix , $\hat{\Sigma}_{t+1}$, and returns, $\hat{\mu}_{t+1}$. As in Callot et al. (2017), we also impose the maximum leverage of the portfolio to be 30% and the maximum weight of one individual asset to be between $[-20\%, 20\%]$. In this setting the problem does not have a closed form solution and it needs to be solved numerically. We also consider as in Brito et al. (2018) a different scenario where short-selling is not allowed, that is, the last restriction becomes $0 \leq w_{i,t+1} \leq 0.20$ and the third one does not apply. We set the transaction cost to 0.1% to both scenarios.

In order to analyze the performance of each forecasting strategy in this exercise, we compute the following statistics[7] in which $H = T - t_0 + 1$ is the number of periods the investor solves the minimisation problem givem by equations (15):

1. Average absolute weight: $\frac{1}{nH}\sum_{t=t_0}^{T}\sum_{i=1}^{n}|\hat{w}_{it}|$;
2. Max weight: $\max_{t_0\leq t\leq T}\max_{1\leq i\leq n}(\hat{w}_{it})$;
3. Min weight: $\min_{t_0\leq t\leq T}\min_{1\leq i\leq n}(\hat{w}_{it})$;
4. Average leverage: $\frac{1}{nH}\sum_{t=t_0}^{T}\sum_{i=1}^{n}|w_{i,t+1}|\,\mathbb{1}(w_{it} < 0)$, the average of the weights of the short position;
5. Portfolio leverage: $\frac{1}{nH}\sum_{t=t_0}^{T}\sum_{i=1}^{n}\mathbb{1}(w_{it} < 0)$;
6. Average turnover: $\frac{1}{nH}\sum_{t=t_0}^{T}\sum_{i=1}^{n}|w_{i,t+1} - w^{hold}_{i,t+1}|$, where $w^{hold}_{i,t+1} = w_{i,t-1}\frac{(1+r_{it-1})}{(1+r_{p,t-1})}$, where r_{it} is the realized return of stock i in time t;
7. Average return of the portfolio: $\mu_p = \frac{1}{H}\sum_{t=t_0}^{T} r_{pt} = \frac{1}{H}\sum_{t=t_0}^{T}\hat{w}_t' r_{it}$;
8. Accumulated return: $\prod_{t=t_0}^{T}(1 + r_{pt})$;
9. Standard deviation: $\sigma_p = \sqrt{\frac{1}{H}\sum_{t=t_0}^{T}(r_{pt} - \mu_p)^2}$;
10. Sharpe ratio: $\frac{\mu_p}{\sigma_p}$;
11. Average diversification ratio: $\frac{1}{H}\sum_{t=t_0}^{T}\frac{\sum_{i=1}^{n}\hat{w}_{it}\sigma_{it}}{\sigma_p}$, where $\sigma_p = \hat{w}_t'\Sigma_{it}\hat{w}_t$ and σ_{it} is the variance of the asset i in time t;

12. Economic value: The economic value is the value of Δ such that for different portfolios p_1 and p_2, we have

$$\sum_{t=t_0}^{T} U(r_{p_1 t}) = \sum_{t=t_0}^{T} U(r_{p_2 t} - \Delta). \tag{16}$$

It represents the maximum return the investor would be willing to sacrifice

each time period in order to capture the performance gains associated with

switching from p_1 to p_2. For instance, p_1 is either no change forecasting strat-

egy or EMWA.

[7]For more details about them, see Callot et al. (2017)

4 Results

4.1 Analysis of the dynamics and models

The advantage of those methods is the possibility of analysing the determinants of the covariance matrices and how they vary over time. Thus, in this section we assess the drivers of the dynamic of the selected models for each selection method and period. Figures 1 and 2 display the size of the models for Lasso and Auto- metrics based models for variance (figure a) and covariances (figure b) equations, respectively. The average size of the selected models is quite stable over the period 2015 to 2017 for both selection methods.

Interestingly, Lasso selected from 55 to 80 variables when the model is a VAR(1) and Var(5) and the size is also similar to diagonal and off-diagonal equations. To VAR(5) the models are slight larger. This result is in line with Callot et al. (2017) that find that Lasso selects model twofold bigger for VAR(20) compared to VAR(1) for daily volatility. Of course, there is a difference of magnitude, but our results indicates the same pattern.

Adaptive Lasso behaves as expected. The number of selected variables for each model is almost half of that with Lasso. It is expected due to Lasso is the first stage of the adaptive Lasso. The sizes are also similar to VAR(1) and VAR(5) models. We will explore the differences in drivers later in this study.

Autometrics selects an even sparse array of parameters, due to the fact that this method satisfies the Oracle property, which is not the case for Lasso and adaLasso. The average model size is around 8 variables when the model is a VAR(1) and 35 when it is a VAR(5). The difference in size for the latter is expected as the significance level, $\alpha = 0.01$ is the same. Autometrics tends to select αK, where K is the number of candidate variables. Theoretically, Autometrics would select 4 variables for VAR(1) and 22 for VAR(5).

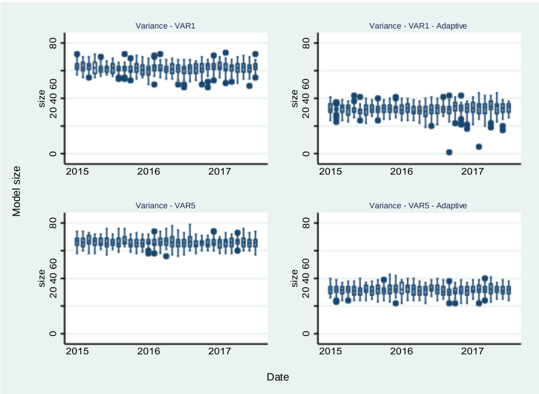
With the inclusion of dummy saturation, the average size for VAR(1) model is stable and around 8 as well, but with a higher dispersion. For VAR(5) model, the average size is closer to 22, although the dispersion is also higher. Due to outliers, more relevant variables may have been excluded. Comparing the dispersion of Autometrics and Lasso, the former presents larger variability in models sizes.

To analyze the stability of the parameters from month to month, figures 3a and 3b plot the frequency of times each variable is selected at period t but not retained at period t+1. Autometrics returns much more stable models, that is, the evolution of the frequency of changes is almost a horizontal line as it can be seen in figures 3a and 3b. The average change in parameters to all equations is below 10%. Interestingly, it is even lower to VAR(5) model and it also presents lower dispersion. For Lasso in figures 4a and 4b, the results are also similar to Callot et al. (2017). The frequency of changes are stable over time but higher than Autometrics models. For Lasso models, they are around 20% for VAR(1) and 5% for VAR(5). The difference between diagonal and off-diagonal equations in terms of changes in parameters is that for the former, the dispersion is higher but the level is lower.

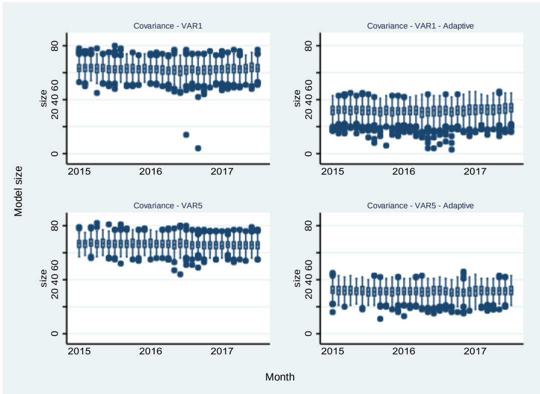
In order to understand better the main driver of the variance and covariance, tables 2 and 3 show the average across estimation windows of the fractions of re-

Figure 1: Models' sizes to Lasso and adaptive Lasso

(a) Variance



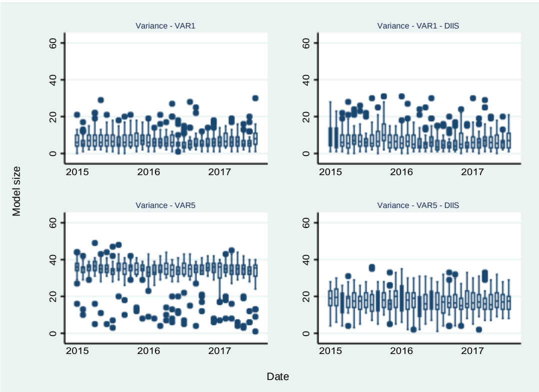
(b) Covariance



Source: The authors.

Figure 2: Models' sizes to Autometrics

(a) Variance

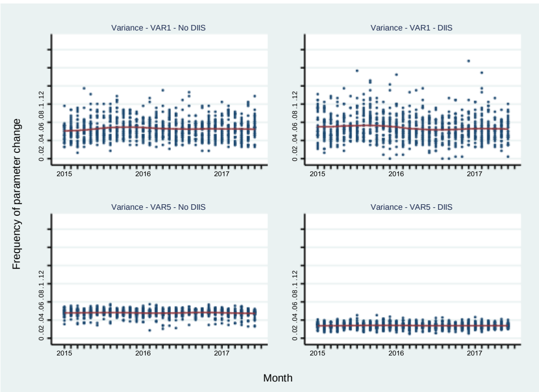


(b) Covariance

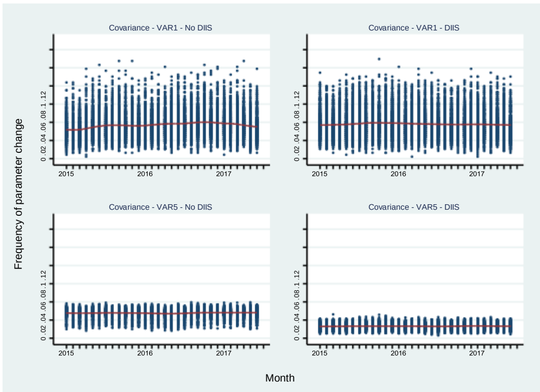
Source: The authors.

Figure 3: Parameter stability analysis to Autometrics models

(a) Variance



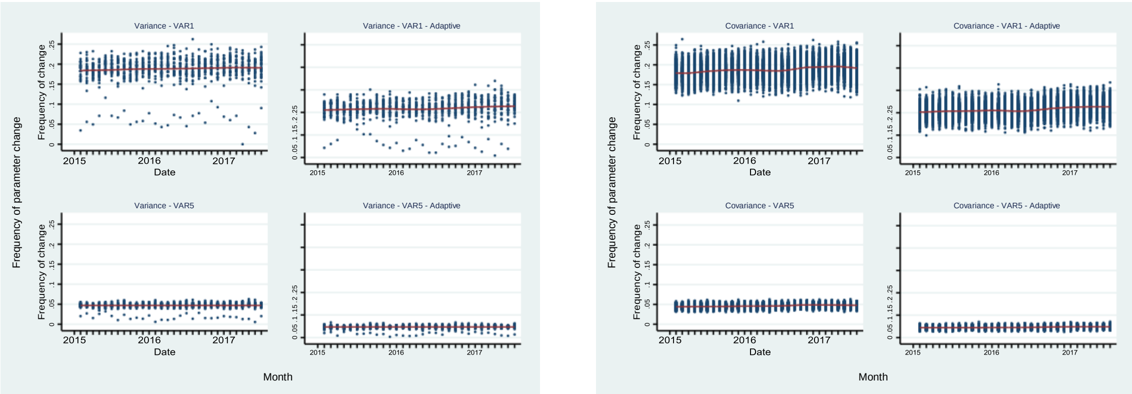
(b) Covariance



Source: The authors.

Figure 4: Parameter stability analysis to Lasso and adaptive Lasso models

(a) Variance (b) Covariance



Source: The authors.

gressors (in rows) from each sector selected to each equation (in columns) from each sector for Autometrics and Lasso VAR(1) models[8]. This number goes from zero to one and we compute as the average of the number of variables from each category in row divided by the number of all candidates of the same type to each sector in the column.

The results are similar for both types of model selection methods and accordingly to what Callot et al. (2017) found for monthly aggregation and Lasso-based forecasts. From table 2 and, to complement from tables A3, A4 and A5 in the appendix, we find that the lagged variance is selected more often for the variance equations. For construction and financial sectors, are more evident from those ta- bles. We see that 16% of the lagged variance of the construction sector is retained in the firms' models from the same sector using Autometrics (table 2). The fraction for financial sector is also 16%. When we use a VAR(5) A4, those fractions decrease to 5% and 4%, respectively.

With the inclusion of dummy variables, the conclusion changes. Tables A3 and A5 report the results. All lagged variances of all sector are almost equally selected. The exception is finance sector which presents 12% of its own past variances as main drivers in a VAR(1) model. This value changes to 2% when a VAR(5) is used. The lagged covariances do not seem to be important to variance equation in any model. Their selection rate are close to zero.

For Lasso based models, we see the same pattern for construction and financial sector though the magnitude of the factions is different. For both, 57% (table 3) of the lagged variance is retained for the diagonal equations for a VAR(1) model and it decreases to 35% and 37%, respectively, when the model is a VAR(5) (table A7). For adaptive Lasso, in tables A6 and A8, those fractions go to below 10%.

For all models, the lagged covariances are little selected to variance equations. This is also found by Callot et al. (2017).

Two points must be highlighted. the first is the difference in fractions between

[8]To save space, the respective tables A3, A4, A5, A6, A7 and A8 for the other models and methods are in appendix.

Table 2: Fraction of regressors (rows) selected to equations (column) to Autometrics based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.16	0.17	0.15	0.00	0.04	0.02	0.00	0.08
Consumer, Cyclical		0.02	0.05	0.00	0.01	0.02	0.01	0.03	0.05
Energy		0.10	0.11	0.06	0.01	0.03	0.01	0.00	0.01
Financial		0.02	0.03	0.18	0.16	0.06	0.01	0.03	0.10
Industry		0.01	0.02	0.04	0.02	0.04	0.07	0.01	0.04
Oil & Mining		0.03	0.01	0.03	0.01	0.03	0.08	0.00	0.01
Services		0.00	0.01	0.02	0.07	0.03	0.02	0.06	0.03
Transport		0.11	0.04	0.08	0.02	0.12	0.06	0.03	0.05
Construction	Lagged covariance	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.00
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
Energy		0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.01
Financial		0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.02
Industry		0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
Transport		0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01
		Covariance							
Construction	Lagged variance	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.02	0.03	0.01	0.02	0.02	0.02	0.03	0.02
Energy		0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Financial		0.07	0.04	0.07	0.06	0.02	0.03	0.03	0.04
Industry		0.03	0.02	0.04	0.03	0.02	0.02	0.02	0.02
Oil & Mining		0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.02
Services		0.02	0.02	0.02	0.03	0.02	0.01	0.02	0.02
Transport		0.04	0.03	0.02	0.03	0.05	0.05	0.06	0.03
Construction	Lagged covariance	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.02
Consumer, Cyclical		0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.02
Energy		0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02
Financial		0.02	0.03	0.02	0.03	0.02	0.03	0.03	0.03
Industry		0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Oil & Mining		0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Services		0.01	0.02	0.02	0.03	0.02	0.02	0.02	0.02

Source: The authors.

Transport	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.02

Source: The authors.

Table 3: Fraction of regressors (rows) selected to equations (column) to Lasso based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.57	0.59	0.27	0.20	0.46	0.21	0.33	0.53
Consumer, Cyclical	Lagged variance	0.38	0.45	0.35	0.30	0.26	0.21	0.50	0.40
Energy	Lagged variance	0.48	0.52	0.46	0.19	0.37	0.32	0.21	0.41
Financial	Lagged variance	0.44	0.31	0.45	0.57	0.29	0.30	0.23	0.28
Industry	Lagged variance	0.33	0.35	0.36	0.32	0.35	0.47	0.33	0.41
Oil & Mining	Lagged variance	0.35	0.23	0.43	0.16	0.36	0.42	0.28	0.20
Services	Lagged variance	0.23	0.23	0.34	0.46	0.39	0.35	0.49	0.43
Transport	Lagged variance	0.34	0.24	0.33	0.27	0.40	0.39	0.37	0.28
Construction	Lagged covariance	0.13	0.11	0.11	0.10	0.10	0.10	0.13	0.08
Consumer, Cyclical	Lagged covariance	0.13	0.12	0.11	0.14	0.13	0.14	0.12	0.13
Energy	Lagged covariance	0.16	0.19	0.15	0.18	0.16	0.17	0.17	0.17
Financial	Lagged covariance	0.06	0.07	0.05	0.06	0.05	0.05	0.06	0.07
Industry	Lagged covariance	0.14	0.14	0.14	0.13	0.14	0.13	0.14	0.13
Oil & Mining	Lagged covariance	0.12	0.11	0.13	0.12	0.12	0.10	0.12	0.11
Services	Lagged covariance	0.08	0.09	0.11	0.11	0.10	0.09	0.10	0.10
Transport	Lagged covariance	0.12	0.11	0.11	0.11	0.11	0.10	0.09	0.12
		Covariance							
Construction	Lagged variance	0.32	0.28	0.24	0.23	0.31	0.27	0.23	0.26
Consumer, Cyclical	Lagged variance	0.38	0.40	0.39	0.38	0.35	0.38	0.40	0.42
Energy	Lagged variance	0.36	0.28	0.34	0.28	0.31	0.30	0.29	0.25
Financial	Lagged variance	0.39	0.37	0.41	0.45	0.27	0.30	0.36	0.34
Industry	Lagged variance	0.32	0.32	0.32	0.34	0.31	0.31	0.32	0.35
Oil & Mining	Lagged variance	0.28	0.26	0.23	0.25	0.29	0.26	0.27	0.25
Services	Lagged variance	0.41	0.40	0.34	0.47	0.35	0.39	0.39	0.42
Transport	Lagged variance	0.34	0.31	0.25	0.33	0.35	0.36	0.36	0.29
Construction	Lagged covariance	0.12	0.12	0.11	0.11	0.12	0.11	0.12	0.11
Consumer, Cyclical	Lagged covariance	0.12	0.13	0.12	0.11	0.12	0.12	0.12	0.12
Energy	Lagged covariance	0.15	0.16	0.17	0.16	0.16	0.17	0.16	0.17
Financial	Lagged covariance	0.07	0.07	0.06	0.06	0.07	0.06	0.06	0.06
Industry	Lagged covariance	0.14	0.14	0.14	0.13	0.14	0.14	0.14	0.14
Oil & Mining	Lagged covariance	0.12	0.11	0.13	0.12	0.12	0.13	0.12	0.12
Services	Lagged covariance	0.10	0.10	0.10	0.10	0.10	0.09	0.10	0.10

Source: The authors.

Transport	0.12	0.12	0.13	0.12	0.12	0.13	0.13	0.13

Source: The authors.

selection methods which may be explained by the difference in size previously discussed. As Lasso does not select through hypothesis testing, there may be more non-significant variables retained, what may be beneficial for prediction as it may reduce the forecasting error. The second is the replacement from lagged variance to lagged covariance when the order of the VAR model increases as reported in tables A6, A7 and A8 in appendix. For instance, for Consumer cyclical sector and VAR(1) in table 2, the relation between fractions of the lagged variance and covariance in the variance equation is 5 while it reduces to 3 for a VAR(5) in table A4.

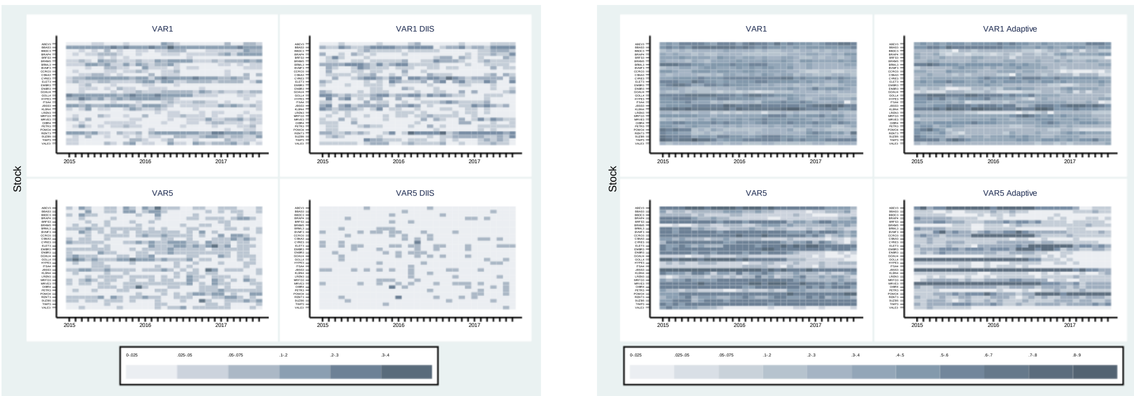
For the covariance equations, the conclusions are also similar between all models and selection methods. The lagged variances are more retained than the lagged covariance, but the difference in fractions is not large. It means that the past variance is relatively more important to the covariances models. We must also note that there are 435 equations, the number of firms is different among sectors and each covariance between two stocks enters in two different rows, so interpretation is not trivial.

In order to explore more the determinants of the variances equations, we plot in figure 5 the percentage of times the lagged variance is retained for each variance equation of each stock in each period. For Autometrics (panel a), the past variance is more retained for VAR(1), especially to Banco do Brazil (BBSA3) and Cyrela (CYRE3) stocks, close to 40%. With the inclusion of dummy saturation, the selection rate reduces, but it is still relatively high to BBSA3. For VAR(5), the past variances are selected less than 10% to most of the stocks.

For Lasso based models, panel b, there is a different conclusion. The past variances are highly selected to most of the periods and stocks, close 90%. Especially for VAR(1), on the top left of panel a. For adaptive LASSO VAR(5) at the end of the period, it seems to change and less lagged variances are selected.

Figure 5: Selection frequency of the diagonal equations

(a) Autometrics (b) Lasso and adaptive Lasso



Source: The authors.

To conclude, both methods show the importance of the lagged variances to the diagonal and off-diagonal equations, besides the differences in the level of retention. It also seems that extreme values may induce the selection of those variables as, using dummy saturation strategy, the selection rate of the past variances reduces to all equations and periods. The differences in size may be explained by the low

significance level we adopt to Autometrics. Increasing the significance level would yield larger models. Autometrics also returns more stable models, that is, the change in selected variables between periods is lower. Besides the opportunity to explore the drivers of the covariance matrix of the selected stocks, this paper aims to compare their forecasting performance what is done in the next sections.

4.2 Forecasting Evaluation

In this section, we forecast the covariance matrix using each selection method and analyse their performance over different horizons: one, two and six months ahead. Table 4 reports the three forecasting error measures, respectively, the average maximal absolute error (AMaxFE), the average median absolute error (AMedFE) and the average g_2 error (L2), for the variance (VAR), covariances (Cov) and all the matrix (All). All values are in relation to the benchmark model, no change. If the number is above one, it means the model is not able to deliver better forecasts than the benchmark.

Table 4 shows that it is very hard to an individual model to beat the benchmark. For all forecasting horizons and error measures the no-change model delivers more accurate predictions. EWMA and Riskmetrics 2006 under-perform by far all models (see table A2 in appendix). Our results are in accordance to Callot et al. (2017) for monthly data.

The equal forecast combination does seem to be very affected by the worse forecasts, specifically EWMA, and it also performs poorly. The weighted combination of forecast, however, is able to beat the benchmark for all horizons and forecast error measures. Specifically, it is more precise for the next month and two periods ahead, $h = \{1, 2\}$, respectively. The level of improvement over the benchmark de- pendents on the forecast error measures. For AMedFE, the forecasts are up to 3% more precise for the whole matrix and covariance equations. Using the AMaxFE, there is a gain of 13% over the no-change forecast for the covariance equation one step ahead ($h = 1$) and 10% for $h = 2$. The variance, the combination outperforms in 4% in the shortest horizon. In the longest, the gain is smaller, 4%. For L2, the weighted combination is more precise up to 9% for the off-diagonal equations. Cava- leri and Ribeiro (2011) also found improvements using different types of forecasting combinations to Ibovespa index volatility.

Differently from Callot et al. (2017), our results show that the covariances are more precisely forecast than the variances. This outcome is robust to horizons and forecast error measures. The improvement seems to be slightly larger using the average maximum forecast error (AMaxFE).

It is also worth noting that among LASSOVAR models, the adaptive LASSO-VAR(1) and LASSOVAR(1) have similar performance, while for VAR(5), LASSO-VAR outperforms adaptive LASSOVAR. As we note in the previous section, it may be explained by the number of variables retained in each model. While retaining more variable reduces the forecasting error, it also increases its variance.

When the model selection is Autometrics, the performance is similar for a VAR(1) with and without dummy saturation. For a VAR(5) model, the inclusion of dummy saturation is beneficial and there is an increase in precision to all fore-

Table 4: Forecast error measures

Model	h	AMedFE			AMAaxFE			L2		
		All	Var	Cov	All	Var	Cov	All	Var	Cov
No change (x 100)	1	0.02	0.03	0.02	0.57	0.57	0.19	1.11	0.71	0.82
Benchmark	2	0.02	0.03	0.02	0.60	0.60	0.18	1.15	0.76	0.82
	6	0.02	0.03	0.02	0.59	0.59	0.16	1.03	0.72	0.70
EWMA	1	15.93	34.14	14.76	9.23	9.23	14.23	14.66	15.40	14.58
(Lambda=0.97)	2	16.05	33.38	14.87	8.83	8.83	14.85	14.16	14.45	14.66
	6	18.55	34.93	17.37	8.91	8.91	16.59	15.97	15.36	17.25
LASSOVAR(1)	1	1.19	1.42	1.19	1.20	1.20	0.94	1.18	1.34	1.01
	2	1.22	1.47	1.21	1.28	1.28	0.97	1.18	1.32	1.06
	6	1.38	1.56	1.38	1.30	1.29	1.09	1.37	1.49	1.19
LASSOVAR(1)	1	1.19	1.44	1.20	1.19	1.19	0.91	1.15	1.28	1.03
Adaptive	2	1.20	1.39	1.20	1.03	1.03	1.01	1.09	1.16	1.08
	6	1.45	1.33	1.46	1.25	1.25	1.16	1.40	1.49	1.24
VAR(1)	1	1.18	1.21	1.18	1.86	1.86	1.04	1.49	1.81	1.09
Autometrics	2	1.30	1.21	1.30	1.99	1.99	1.24	1.56	1.81	1.21
	6	1.74	1.17	1.77	2.42	2.42	1.62	2.10	2.33	1.64
VAR(1)	1	1.19	1.36	1.19	2.07	2.07	1.05	1.71	2.13	1.14
Com IIS+SIS+DIIS	2	1.26	1.38	1.25	3.75	3.75	1.19	2.54	3.42	1.20
Autometrics	6	1.46	1.36	1.44	4.08	4.08	1.49	2.98	3.83	1.38
LASSOVAR(5)	1	1.07	1.28	1.07	1.03	1.03	0.96	1.07	1.15	0.99
	2	1.12	1.17	1.13	0.94	0.93	1.00	1.01	1.01	1.02
	6	1.18	1.43	1.17	1.02	1.01	0.93	1.14	1.19	1.04
LASSOVAR(5)	1	1.11	1.19	1.11	1.20	1.19	1.00	1.15	1.28	1.03
Adaptive	2	1.15	1.20	1.16	1.13	1.13	1.00	1.11	1.16	1.05
	6	1.21	1.43	1.21	2.26	2.26	0.99	1.79	2.15	1.07
VAR(5)	1	1.35	1.35	1.35	1.47	1.47	1.18	1.41	1.56	1.25
Autometrics	2	1.42	1.53	1.42	3.44	3.44	1.30	2.41	3.18	1.33
	6	1.71	1.47	1.74	5.14	5.14	1.60	3.75	4.91	1.61
VAR(5)	1	1.21	1.45	1.20	1.39	1.39	1.14	1.32	1.50	1.16
Com IIS+SIS+DIIS	2	1.24	1.27	1.24	1.68	1.68	1.14	1.48	1.75	1.17
Autometrics	6	1.31	1.28	1.32	4.37	4.37	1.30	3.04	4.00	1.24
Combination	1	2.02	3.96	1.93	1.40	1.40	1.72	1.90	2.05	1.81
Equal	2	2.04	3.98	1.95	1.63	1.63	1.84	2.00	2.20	1.88
	6	2.42	4.53	2.30	1.91	1.91	2.21	2.50	2.75	2.23
Combination	1	0.97	1.37	0.97	0.96	0.96	0.87	0.98	1.05	0.91
Weighted	2	0.98	1.31	0.97	1.22	1.22	0.90	1.08	1.24	0.93
	6	1.01	1.36	0.98	1.28	1.28	0.96	1.28	1.50	0.95

Note: The number for No-change are in absolute values and the others are relative to the benchmark. All: All the covariance matrix; Var: only the variance; Cov: only the covariance. We forecast h months ahead where h = 1, 2, 6.
Source: The authors.

casting horizons, with some exceptions, for example, variance forecast 1 step ahead with increases from 35

Comparing Autometrics and LASSO, we note that for a VAR(1), the performance is quite similar while LASSO and adaptive LASSO outperform Autometrics we use VAR(5) model. It is worth noting that the tendency of VAR(5) forecast better than the VAR(1) may be a consequence of the long memory characteristics presented in volatility series (Granger et al., 2000, for instance).

In order to check the statistically differences in the models' forecast, we calculate the Model Confidence Set (Hansen et al., 2011) and Giacomini and Rossi (2010)'s conditional test. The results are in tables 5 and 6, respectively.

Table 5: Model Confidence Set for all equations

Models	AMedFE			AMaxFE			L2		
	h=1	h=2	h=6	h=1	h=2	h=6	h=1	h=2	h=6
No change	0.48	0.53	1.00	0.77	0.51	1.00	0.76	1.00	1.00
EWMA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LASSOVAR(1)	0.00	0.00	0.00	0.12	0.25	0.41	0.03	0.11	0.02
AdaLASSOVAR(1)	0.00	0.00	0.00	0.29	0.50	0.62	0.03	0.17	0.01
Autometrics VAR(1)	0.01	0.00	0.00	0.47	0.30	0.14	0.23	0.15	0.00
Autometrics VAR(1) DIIS	0.00	0.00	0.00	0.47	0.26	0.05	0.23	0.17	0.03
LASSOVAR(5)	0.30	0.08	0.03	0.69	1.00	0.93	0.27	0.83	0.41
AdaLASSOVAR(5)	0.01	0.03	0.00	0.47	0.30	0.62	0.13	0.11	0.36
Autometrics VAR(5)	0.00	0.00	0.00	0.04	0.04	0.02	0.00	0.02	0.00
Autometrics VAR(5) DIIS	0.00	0.00	0.00	0.12	0.26	0.25	0.01	0.04	0.08
Combination equal	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Combination weight	1.00	1.00	0.80	1.00	0.50	0.62	1.00	0.71	0.41

Note: The table displays the p-values of the MCS. We forecast h = 1, 2, 6 months ahead.
Source: The authors.

The MCS depends on the error measure, but still either the weighted combination or the no-change models is the winner. One exception is LASSOVAR(5) for AMaxFE and h = 2. For the average median forecasting error (AMedFE) the weighted combination delivers the most precise predictions for h = /1, 2/. For the h = 6, the final set is composed of the former and the no-change forecast.

The model confidence sets using the average maximal forecasting error (AMaxFE) have more elements than using the previous measure. The ranking is also different depending on the horizon. For one month ahead, the best predictions are from the weighted combination approach followed no-change forecast. For the two months ahead, the winner is LASSOVAR(5) followed by the no-change forecast and weighted combination. For h = 6, the benchmark comes first and LASSOVAR(5) delivers the second most accurate predictions, while the weighted combination ranks in third. Callot et al. (2017) also found Lasso based prediction are more precise in longer horizons and forecasting error measure. This measure is important as they derived the theoretical upper limit of the covariance matrix forecasting error.

Using the last error measure, L2, the weighted combination shows better predictions for h = 1, while the benchmark is in the second. For h = /2, 6/, those models change the position.

In table 6, we report Giacomini and Rossi (2010)'s conditional test. It is a one to one test, so we show the result to the null of equal predictive ability between the benchmark (no-change forecast) and the alternative model. P-values are reported next to the sign that means the benchmark is more precise if negative and the opposite if positive. The results are not very different from the previous one. At 5% of significance level, the benchmark returns better or equal forecast compared to alternative methods. The weighted combination forecast the only strategy that delivers statistically equal prediction at 5% significance level.

The conclusions confirm that it is not easy for individual models to beat a random walk model. However, there are gains in forecasting combination for some forecasting horizons. The gains for combining the predictions goes from 1 to 13% over the benchmark. In the next section, we calculate the economic value of each model applying the forecasts to form investment portfolios each month.

5 Portfolio Selection

In this section, we use the covariance matrix to construct investment portfolios. There are two setting based on Fleming et al. (2001, 2003), Callot et al. (2017) and Brito et al. (2018) where an agent construct every month his investment portfolio minimising its variance. We consider a setting with and without short-selling constraint. Comparing the forecasts through this exercise has two advantages: the first that it does not depend on the true covariance matrix as we can see the accumulated return of each investment strategy; the second is that one may argue that the error measures used in the previous section is too aggregate, specially taking into account the number of unique entries in the covariance matrix.

Table 7 reports the results when short-selling is possible. As the objective is to minimize the portfolio's variance, we first compare their standard deviation. The no-change forecasts returns a portfolio with 4.25% of standard deviation. Just three models present slight smaller standard deviation than the benchmark, Au- tometrics VAR(5) with dummy saturation, 3.78%, LASSOVAR(1) and adaptive LASSOVAR(1), 4.14% and 4.01%, respectively. However, alternative methods have very competitive accumulated returns. Autometrics VAR(1) with dummy satura- tion reached 76.63% while the benchmark is 44.08% and the second highest return is from adaptive LASSOVAR(1), 61.77%.

The Sharpe-ratios show that Adaptive LASSOVAR(1) outperform all the other models, followed by LASSOVAR(1) and Autometrics VAR(1) with DIIS. Again, almost all models delivered higher Sharpe-ratio than our benchmark.

Curiously, the weighted combination forecast delivers worse accumulated returns and a slightly lower Sharpe-ratio and higher standard deviation than the benchmark.

The descriptive statistics of the portfolios vary little among portfolios. The average absolute weight goes from 3.98% to 4.77%. Interestingly, both the limits are the no-change and EWMA forecasts. It is also worth noting that the proportion of leverage portfolio is slight above 20% to all models but EWMA, which increase to 30%. Its leverage with its low average turnover explain its relatively high accumulated return and standard deviation. In Brito et al. (2018), EWMA also returned

Table 6: Giacomini and White (2006)'s test: comparing automatic selection based models to the benchmark

Models	AMedFE			AMaxFE			L2		
	h=1	h=2	h=6	h=1	h=2	h=6	h=1	h=2	h=6
No change									
EWMA	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -
LASSOVAR(1)	0.03 -	0.02 -	0.00 -	0.38 -	0.09 -	0.46 -	0.32 -	0.20 -	0.00 -
AdaLASSOVAR(1)	0.04 -	0.01 -	0.00 -	0.28 -	0.64 -	0.42 -	0.22 -	0.12 -	0.00 -
Autometrics VAR(1)	0.07 -	0.00 -	0.00 -	0.37 -	0.04 -	0.19 -	0.29 -	0.02 -	0.00 -
Autometrics VAR(1) DIIS	0.02 -	0.00 -	0.00 -	0.44 -	0.14 -	0.02 -	0.34 -	0.13 -	0.00 -
LASSOVAR(5)	0.38 -	0.20 -	0.00 -	0.21 -	0.07 +	0.05 -	0.59 -	0.42 -	0.42 -
AdaLASSOVAR(5)	0.18 -	0.04 -	0.00 -	0.43 -	0.49 -	0.06 -	0.14 -	0.50 -	0.52 -
Autometrics VAR(5)	0.01 -	0.00 -	0.00 -	0.19 -	0.04 -	0.00 -	0.04 -	0.02 -	0.00 -
Autometrics VAR(5) DIIS	0.01 -	0.00 -	0.00 -	0.07 -	0.18 -	0.02 -	0.02 -	0.06 -	0.00 -
Combination equal	0.00 -	0.00 -	0.00 -	0.06 -	0.00 -	0.00 -	0.00 -	0.00 -	0.00 -
Combination weight	0.49 +	0.30 +	0.11 -	0.26 +	0.17 -	0.07 -	0.73 +	0.25 -	0.26 -

Note: The table displays the p-values of Giacomini and White (2006) which has the null of equality of predictive power for forecast *h* steps ahead. The sign (-) means that the benchmark has smaller forecasting error than the respective model in the row. The sign (+) means the no-change forecast has greater forecasting error.
Source: The authors.

Table 7: Portfolio selection with short-selling

Model	VAR(1)				VAR(5)				Combination		No Change	EWMA
Estimator	Lasso		Auto		Lasso		Auto					
Statistic		Adap.		DIIS		Adap.		DIIS	Equal	Weight		
Av. abs. weight (%)	4.04	3.97	3.99	4.03	4.14	4.11	4.07	4.05	4.33	4.06	3.98	4.77
Max weight (%)	13.20	13.88	13.80	14.97	12.85	12.67	15.17	13.67	14.87	15.07	13.21	19.40
Min weight (%)	-7.45	-7.11	-8.50	-10.50	-8.03	-6.73	-9.24	-7.92	-7.18	-7.73	-6.69	-7.45
Average leverage (%)	-10.59	-9.59	-9.81	-10.40	-12.10	-11.63	-11.04	-10.75	-14.96	-10.92	-9.74	-21.60
Proportion of leverage (%)	23.16	21.75	21.40	22.98	25.44	25.09	23.86	24.04	29.65	24.04	21.23	31.58
Average turnover (%)	3.00	2.87	3.12	3.34	3.39	3.18	3.66	3.45	1.67	2.52	2.86	1.13
Average % return	2.51	2.64	1.47	3.17	2.34	2.10	1.99	2.17	2.10	1.88	2.02	2.28
Accumulated % return	57.91	61.77	29.53	76.63	51.92	45.58	43.58	47.25	45.80	39.98	44.08	50.94
Standard deviation	4.14	4.01	4.55	5.42	4.99	4.60	3.78	4.97	4.51	4.43	4.25	4.35
Sharpe ratio	0.61	0.66	0.32	0.59	0.47	0.46	0.53	0.44	0.47	0.42	0.48	0.52
Diversification ratio	2.10	2.10	2.14	2.10	2.12	2.10	2.03	2.06	2.13	2.16	2.14	2.07
Economic Value ($\gamma = 1$)												
No change	6.08	7.74	-6.62	13.95	3.46	0.68	-0.21	1.40	0.77	-1.84		
EWMA	2.95	4.57	-9.40	10.61	0.40	-2.30	-3.17	-1.60	-2.21	-4.75		
Economic Value ($\gamma = 5$)												
No change	6.35	8.31	-7.25	10.69	1.66	-0.13	0.77	-0.34	0.19	-2.23		
EWMA	3.45	5.37	-9.82	7.63	-1.15	-2.88	-1.98	-3.10	-2.56	-4.92		
Economic Value ($\gamma = 10$)												
No change	6.95	9.58	-8.69	2.90	-2.54	-1.98	2.95	-4.39	-1.13	-5.32		
EWMA	4.61	7.21	-10.81	0.22	-4.92	-4.26	0.76	-6.72	-3.39	0.00		

Source: The authors.

more leveraged portfolios, but in a different context, with no limits in the short-selling.

In the last part of the table, there is the annualized return an investor is willing to pay to change from the benchmark strategy to an alternative approach. Autometrics VAR(1) with DIIS presents the highest economic value, 13.95%, followed by Adaptive LASSOVAR(1), 7.74%, and LASSOVAR(1), 6.08%. As the risk-aversion increases, Adaptive LASSOVAR(1) becomes better as it delivers lowers standard deviations. When $\gamma = 10$, its economic value becomes 9.58% compared to 2.90% of the Autometrics VAR(1) with DIIS.

When short-selling is not permitted, table 8 reports the results. The most leveraged portfolio formed by EWMA covariance matrix is the one that most suffered. It increased its standard deviation and reduces its return. Its Sharpe-ratio is not the same as the benchmark. With smaller magnitude, the benchmark also returned more riskier portfolio with lower return.

Comparing the standard deviation among all portfolios, Autometrics VAR(1) delivered the less risky one, 3.96%, followed by the benchmark, 4.29%. When the metrics is accumulated return, Autometrics VAR(1) with DIIS is also the winner. It reduces just 3 percentage point, from 76.63% to 73.08%. It also returned the highest Sharpe-ratio, 0.66.

Looking at the last bottom panel of table 8, it is possible to see that Autometrics VAR(1) DIIS delivers 13.94% of economic value. As risk aversion increases, it decreases but as a slower rate due to the high Sharpe-ratio.

Apart from Autometrics VAR(1), which delivered the lowest accumulated return, standard deviation and Sharpe-ratio, and equally weighted combination of forecasts, all alternative models have positive economic value for all values of γ.

Summing up this section, the results show that in the context of portfolio investment, it is possible to outperform the benchmark, although not all of model are able to do it. Moreover, the benchmark is more competitive when short-selling is not a constraint to the agent. However, Autometrics VAR(5) with dummy saturation outperforms all its competitors in both settings in terms of accumulated return. Adaptive LASSOVAR(1) stands out when short-selling is allowed and the investor is more risk-averse.

6 Conclusion

Modelling and forecasting the covariance matrix have become very important in economics. It is an essencial part in the modern portfolio allocation theory, for instance, (Markowitz, 1952). It is also important for risk management, option pricing, security regulations and hedging strategies. With the development of financial markets, the number of stocks and assets available has increased fast. Consequently, it poses a challenge to standard methods broadly used in the literature (Baba et al., 1990; Engle, 2002). They fail to deliver precise estimates of covariance matrix due to the curse of dimensionality (Callot et al., 2017).

New methods have been developed to estimate and model its dynamic. Callot et al. (2017) develops a vector autoregression model in which each individual entry of the covariance matrix is one equation, yeilding a highly dimensional system. They

Table 8: Portfolio selection without short-selling

Model	VAR(1)				VAR(5)							
Estimator	Lasso		Auto		Lasso		Auto		Combination		No Change	EWMA
Statistic		Adap.		DIIS		Adap.		DIIS	Equal	Weight		
Av. abs. weight (%)	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33	3.33
Max weight (%)	19.82	19.78	19.88	19.65	19.50	19.40	19.70	19.78	19.62	19.69	19.66	19.95
Min weight (%)	0.02	0.02	0.01	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.00
Average leverage (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Proportion of leverage (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Average turnover (%)	2.87	2.91	2.66	3.04	2.80	2.77	3.22	2.92	1.42	2.33	2.80	0.88
Average % return	2.36	2.30	1.66	3.03	2.45	2.28	2.11	2.37	2.00	2.28	1.92	2.03
Accumulated % return	52.86	51.07	35.00	73.08	55.71	50.91	46.22	52.80	43.15	50.79	41.25	43.80
Standard deviation	4.74	4.75	3.96	4.59	4.43	4.40	4.43	4.93	4.59	4.43	4.29	4.65
Sharpe ratio	0.50	0.48	0.42	0.66	0.55	0.52	0.48	0.48	0.44	0.51	0.45	0.44
Diversification ratio	2.11	2.13	2.13	2.12	2.11	2.10	2.07	2.11	2.12	2.17	2.13	2.08
Economic Value ($\gamma = 1$)												
No change	5.19	4.40	-2.85	13.94	6.48	4.36	2.26	5.16	0.87	4.29		
EWMA	3.98	3.20	-3.97	12.64	5.25	3.16	1.08	3.96	-0.30	3.09		
Economic Value ($\gamma = 5$)												
No change	4.10	3.29	-2.16	13.17	6.16	4.12	1.94	3.58	0.19	3.97		
EWMA	3.76	2.95	-2.48	12.80	5.81	3.77	1.60	3.24	-0.14	3.63		
Economic Value ($\gamma = 10$)												
No change	1.63	0.78	-0.63	11.44	5.44	3.57	1.22	-0.03	-1.34	3.25		
EWMA	3.24	2.38	0.83	13.16	7.05	5.14	2.77	1.59	0.20	4.83		

show that it is possible to find a sparse matrix of parameter using Lasso (Tibshirani, 1996; Zou, 2006). This paper aims to use a different selection method to model this system. We apply the general to specific approach developed by Hendry and Krolzig (1999) and Doornik (2009), named Autometrics. Thus, the objective of the this study is to compare the forecasting performance of those methods and the combination of forecasts.

To do so, a database is composed with 30 assets which yields 465 equations and up to 1,081,590 parameters considering a VAR(5) model. The stocks are from Brazilian stock market, B3, for the period July/2009 to December/2017. We use three forecasting error measures and compute the model confidence set (Hansen et al., 2011). Furthermore, we implement a real exercise where an investor needs to form his portfolio every month and compare the investment performance using each forecast strategy.

The results shows that the final selected model is very sparse. Out of 465 possible regressor for each equation in a VAR(1), Autometrics selects on average eight and Lasso 60. The difference in size can be explained by the significance value we assume to Autometrics. Increasing the value, more variables are retained. Both methods deliver quite stable models over the period, in terms of size and stability of parameters. When analyzing the determinants of covariance matrices' dynamics, the past variances seem to be more important for the variance and for the covariance equations. We do not find that one important sector is the most relevant to drive the covariance matrices. Furthermore, there is a slight evidence that the past variances of some sectors are the main drivers of their own volatility, as expected.

The forecasting analysis indicates that it is not easy for an individual model to beat the benchmark of no-change in volatility, that, it, a random walk, but a weighted combination of forecasts is able to increase precision up to 13%. In the portfolio allocation exercise, the results show Autometrics VAR(1) with dummy saturation delivers the highest accumulated return with and without shot-selling constraint. We compute that an investor is willing to pay around to 14% to change from the benchmark strategy to Autometrics VAR(1) with dummy saturation. How- ever, when the investor to very risk-averse, he prefers adaptive LASSOVAR(1) based forecasts and he is willing to pay 10% over 2.9% to Autometrics VAR(1) with dummy saturation.

To conclude, this study shows that there are economic gains using automatic model selection techniques to reduce the dimensionality of systems when modelling and forecasting the covariance matrix for portfolio selection. However, Taylor (2013) finds that the gains vary over time, risk preferences, economic conditions and utility function forms. Future reseaches could explore differences in those aspects. It would be interesting to define sub-periods with high volatility to, specifically, test the performance of Autometrics with impulse dummy saturation. Other authors, such as, Ericsson (2012), find that it increases forecasting accuracy, but not in the context of covariance systems. Other extension is to estimate the covariances using different methodologies (Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2004; Lunde et al., 2016), especially for logarithmic covariances as discussed in Sucarrat and Escribano (2010).

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of financial economics 61* (1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71* (2), 579–625.

Baba, Y., R. F. Engle, D. F. Kraft, and K. F. Kroner (1990). Multivariate si- multaneous generalized arch. *Manuscript, University of California, San Diego, Department of Economics.*

Bai, J., S. Shi, et al. (2011). Estimating high dimensional covariance matrices and its applications. *Annals of Economics and Finance 12* (2), 199–215.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Measuring the impact of jumps in multivariate price processes using bipower covariation. Technical report, Discussion paper, Nuffield College, Oxford University.

Bates, J. M. and C. W. Granger (1969). The combination of forecasts. *Or* , 451–468.

Bauer, G. H. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics 160* (1), 93–101.

Bickel, P. J., E. Levina, et al. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics 36* (1), 199–227.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, 498–505.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy 96* (1), 116– 131.

Boudt, K., J. Danielsson, and S. Laurent (2013). Robust forecasting of dynamic conditional correlation garch models. *International Journal of Forecasting 29*(2), 244–257.

Brito, D., M. C. Medeiros, and R. Ribeiro (2018). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage.

Bucci, A. (2018). Forecasting realized volatility: A review. *Journal of Advanced Studies in Finance 8* (2), 94–138.

Callot, L., A. Kock, and M. Medeiros (2017). Estimation and forecasting of large realized covariance matrices and portfolio choice. *Journal of Business and Eco- nomics Statistics 32* (1), 140–158.

Callot, L. A. and A. B. Kock (2014). Oracle efficient estimation and forecasting with the adaptive lasso and the adaptive group lasso in vector autoregressions. *Essays in Nonlinear Time Series Econometrics*, 238–268.

Campos, J., D. F. Hendry, and H.-M. Krolzig (2003). Consistent model selection by an automatic gets approach. *Oxford Bulletin of Economics and Statistics 65*, 803–819.

Castle, J. L. and D. F. Hendry (2010). A tale of 3 cities: Model selection in over-, exact, and under-specified equations.

Castle, J. L., X. Qin, and W. Robert Reed (2013). Using model selection algorithms to obtain reliable coefficient estimates. *Journal of Economic Surveys 27*(2), 269–296.

Cavaleri, R. and E. P. Ribeiro (2011). Combinação de previsões de volatilidade: um estudo. *Revista Economia 12* (2), 239–261.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26* (6), 922–947.

Chiu, T. Y., T. Leonard, and K.-W. Tsui (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association 91*(433), 198–210.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7* (2), 174–196.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economics Statistics 13.*

Diebold, F. X. and P. Pauly (1987). Structural change and the combination of forecasts. *Journal of Forecasting 6* (1), 21–40.

Doornik, J. A. (2009). Autometrics. In *In Honour of David F. Hendry*. Citeseer.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20* (3), 339–350.

Engle, R. (2009). *Anticipating correlations: a new paradigm for risk management*. Princeton University Press.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric theory 11* (1), 122–150.

Engle, R. F., O. Ledoit, and M. Wolf (2017). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 1–13.

Epprecht, C., D. Guegan, Á. Veiga, and J. Correa da Rosa (2019). Variable selection and forecasting via automated methods for linear models: Lasso/adalasso and autometrics. *Communications in Statistics-Simulation and Computation*, 1–20.

Ericsson, N. R. (2012). Detecting crises, jumps, and changes in regime. *draft, Board of Governors of the Federal Reserve System, Washington, DC, November.*

Fan, J., Y. Liao, and H. Liu (2016). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal 19*(1), C1–C32.

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresh- olding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75* (4), 603–680.

Fleming, J., C. Kirby, and B. Ostdiek (2001). The economic value of volatility timing. *The Journal of Finance 56* (1), 329–352.

Fleming, J., C. Kirby, and B. Ostdiek (2003). The economic value of volatility timing using “realized” volatility. *Journal of Financial Economics 67* (3), 473–509.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25* (4), 595–620.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74* (6), 1545–1578.

Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012). The conditional autoregressive wishart model for multivariate stock market volatility. *Journal of Econometrics 167* (1), 211–223.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181.

Granger, C. W. and R. Ramanathan (1984). Improved methods of combining fore-

casts. *Journal of Forecasting 3* (2), 197–204.

Granger, C. W., S. Spear, and Z. Ding (2000). Stylized facts on the temporal and distributional properties of absolute returns: An update. In *Statistics and Finance: An Interface*, pp. 97–120. World Scientific.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37* (3), 424–438.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79* (2), 453–497.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting 13*(2), 281–291.

Hautsch, N., L. M. Kyj, and R. C. Oomen (2012). A blocking and regularization approach to high-dimensional realized covariance estimation. *Journal of Applied Econometrics 27* (4), 625–645.

Hendry, D. F. (1995). *Dynamic econometrics*. Oxford University Press on Demand.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal 7* (1), 1–31.

Hendry, D. F. and H.-M. Krolzig (1999). Improving on 'data mining reconsidered'by kd hoover and sj perez. *The econometrics journal 2* (2), 202–219.

Hendry, D. F. and H.-M. Krolzig (2005). The properties of automatic gets modelling. *The Economic Journal 115* (502), C32–C61.

Hlouskova, J., K. Schmidheiny, and M. Wagner (2009). Multistep predictions for multivariate garch models: Closed form solution and the value for portfolio management. *Journal of Empirical Finance 16*(2), 330–336.

Hoover, K. D. and S. J. Perez (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The econometrics journal 2* (2), 167–191.

Hotta, L. K. and C. Truc´ıos (2018). Inference in (m) garch models in the presence of additive outliers: Specification, estimation, and prediction. In *Advances in Mathematics and Applications*, pp. 179–202. Springer.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. *Castle, and Shephard (2009)*, 1–36.

Kock, A. B. and T. Teräsvirta (2014). Forecasting performances of three automated modelling techniques during the economic crisis 2007–2009. *International Journal of Forecasting 30* (3), 616–631.

Ledoit, O. and M. Wolf (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Stud- ies 30* (12), 4349–4388.

Lunde, A., N. Shephard, and K. Sheppard (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to port- folio choice. *Journal of Business & Economic Statistics 34*(4), 504–518.

Markowitz, H. (1952). Portfolio selection. *The journal of finance 7* (1), 77–91.

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics 99* (2), 195–223.

Merton, R. C. (1980). On estimating the expected return on the market: An ex- ploratory investigation. *Journal of financial economics 8* (4), 323–361.

Morgan, J. et al. (1996). Riskmetrics technical document.

Newbold, P. and C. W. Granger (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General),* 131–165.

Newey, W. K. and K. D. West (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review* , 777–787.

Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2017). Fitting vast dimensional time-varying covariance models.

Samuels, J. D. and R. M. Sekkel (2017). Model confidence sets and forecast combination. *International Journal of Forecasting 33* (1), 48 – 60.

Sheppard, K. (2013). Mfe toolbox. *University of Oxford. Available at: https://www.kevinsheppard. com/MFE Toolbox# Last Updated* .

Sucarrat, G. and A. Escribano (2010). The power log-garch model.

Swanson, N. R. and H. White (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics 13*(3), 265–275.

Swanson, N. R. and H. White (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics 79* (4), 540–550.

Taylor, N. (2013). The economic value of volatility forecasts: A conditional ap- proach. *Journal of Financial Econometrics 12* (3), 433–478.

Teräsvirta, T., D. van Dijky, and M. C. Medeirosz (2003). Smooth transition autoregressions, neural networks, and linear models in forecasting macroeconomic time series: A re-examination.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological),* 267–288.

Truc´ıos, C., L. K. Hotta, and P. L. V. Pereira. On the robustness of the principal volatility components. *Journal of Empirical Finance 52*, 201–219.

Truc´ıos, C., L. K. Hotta, and E. Ruiz (2018). Robust bootstrap densities for dy- namic conditional correlations: implications for portfolio selection and value-at- risk. *Journal of Statistical Computation and Simulation 88*(10), 1976–2000.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society* , 1067–1084.

Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *Journal of Machine learning research 7* (Nov), 2541–2563.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association 101* (476), 1418–1429.

Zumbach, G. O. (2007). The riskmetrics 2006 methodology. *Available at SSRN 1420185* .

Appendix

Table A1: List of stock

Order	Code	Name	Sector
1	ABEV3	Ambev S/A	Industry
2	BBAS3	Brasil	Financial
3	BBDC3	Bradesco	Financial
4	BRAP4	Bradespar	Services
5	BRFS3	BRF SA	Consumer, cyclical
6	BRKM5	Braskem	Industry
7	BRML3	BR Malls Par	Services
8	BVMF3	B3	Financial
9	CCRO3	CCR SA	Transport
10	CSNA3	Sid Nacional	Industry
11	CYRE3	Cyrela Realt	Construction
12	ELET3	Eletrobras	Energy
13	EMBR3	Embraer	Industry
14	ENBR3	Energias BR	Energy
15	GOAU4	Gerdau Met	Industry
16	GOLL4	Gol	Transport
17	HYPE3	Hypermarcas	Industry
18	ITSA4	Itausa	Services
19	JBSS3	JBS	Consumer, cyclical
20	KLBN4	Klabin S/A	Industry
21	LREN3	Lojas Renner	Consumer, cyclical
22	MRFG3	Marfrig	Consumer, cyclical
23	MRVE3	MRV	Construction
24	OIBR4	Oi	Mining and Oil
25	PETR3	Petrobras	Mining and Oil
26	POMO4	Marcopolo	Industry
27	RENT3	Localiza	Transport
28	SUZB5	Suzano Papel	Industry
29	TIMP3	Tim Part S/A	Services
30	VALE3	Vale	Mining and Oil

Source: The authors.

Table A2: Comparison with an alternative Benchmark

Model	h	AMedFE			AMAaxFE			L2		
		All	Var	Cov	All	Var	Cov	All	Var	Cov
No change (x 100)	1	0.02	0.03	0.02	0.57	0.57	0.19	1.11	0.71	0.82
Benchmark	2	0.02	0.03	0.02	0.60	0.60	0.18	1.15	0.76	0.82
	6	0.02	0.03	0.02	0.59	0.59	0.16	1.03	0.72	0.70
EWMA	1	15.93	34.14	14.76	9.23	9.23	14.23	14.66	15.40	14.58
(Lambda=0.97)	2	16.05	33.38	14.87	8.83	8.83	14.85	14.16	14.45	14.66
	6	18.55	34.93	17.37	8.91	8.91	16.59	15.97	15.36	17.25
Riskmetrics 2006	1	19.62	37.09	18.56	11.64	11.64	19.55	18.38	18.57	18.83
	2	17.87	32.89	16.85	10.08	10.08	18.25	16.00	15.76	17.02
	6	14.73	25.91	13.90	7.12	7.12	14.08	12.76	11.98	14.03

Source: The authors.

Table A3: Fraction of regressors (rows) selected to equations (column) to Autometrics VAR(1) with DIIS models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
					Variance				
Construction	Lagged variance	0.08	0.13	0.03	0.04	0.03	0.03	0.00	0.07
Consumer, Cyclical		0.02	0.05	0.06	0.00	0.02	0.01	0.03	0.03
Energy		0.08	0.08	0.02	0.01	0.01	0.00	0.00	0.02
Financial		0.03	0.01	0.10	0.12	0.04	0.02	0.01	0.04
Industry		0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.02
Oil & Mining		0.01	0.02	0.04	0.01	0.03	0.03	0.01	0.01
Services		0.01	0.01	0.03	0.04	0.03	0.02	0.05	0.03
Transport		0.07	0.02	0.05	0.01	0.10	0.03	0.03	0.03
Construction	Lagged covariance	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.01
Energy		0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.02
Financial		0.01	0.02	0.03	0.01	0.01	0.01	0.02	0.02
Industry		0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01
Oil & Mining		0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01
Services		0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.01
Transport		0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01
					Covariance				
Construction	Lagged variance	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.02
Consumer, Cyclical		0.02	0.02	0.01	0.01	0.02	0.02	0.02	0.02
Energy		0.02	0.02	0.03	0.01	0.02	0.02	0.02	0.01
Financial		0.04	0.03	0.06	0.04	0.02	0.02	0.03	0.02
Industry		0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
Oil & Mining		0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.02
Services		0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.02
Transport		0.03	0.02	0.02	0.02	0.03	0.04	0.03	0.02
Construction	Lagged covariance	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Consumer, Cyclical		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Energy		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Financial		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Industry		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Oil & Mining		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Services		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Source: The authors.

Transport	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Source: The authors.

Table A4: Fraction of regressors (rows) selected to equations (column) to Autometrics VAR(5) models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.05	0.03	0.05	0.03	0.02	0.02	0.04	0.05
Consumer, Cyclical		0.03	0.03	0.02	0.02	0.03	0.02	0.02	0.03
Energy		0.05	0.03	0.07	0.03	0.03	0.03	0.03	0.02
Financial		0.03	0.02	0.03	0.04	0.02	0.03	0.02	0.03
Industry		0.02	0.02	0.03	0.02	0.02	0.03	0.02	0.02
Oil & Mining		0.02	0.02	0.03	0.02	0.02	0.03	0.02	0.01
Services		0.01	0.02	0.02	0.02	0.02	0.03	0.02	0.02
Transport		0.05	0.02	0.05	0.03	0.04	0.03	0.04	0.03
Construction	Lagged covariance	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Transport		0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
		Covariance							
Construction	Lagged variance	0.04	0.03	0.04	0.04	0.02	0.03	0.03	0.04
Consumer, Cyclical		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Energy		0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Financial		0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02
Industry		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Oil & Mining		0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.03
Services		0.02	0.03	0.02	0.03	0.02	0.02	0.02	0.02
Transport		0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.03
Construction	Lagged covariance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.02
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Transport	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Table A5: Fraction of regressors (rows) selected to equations (column) to Autometrics VAR(5) with DIIS based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.03	0.01	0.04	0.01	0.01	0.01	0.01	0.02
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.01	0.02	0.04	0.02	0.02	0.00	0.03	0.02
Financial		0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
Transport		0.02	0.01	0.01	0.02	0.01	0.00	0.01	0.02
Construction	Lagged covariance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Financial		0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Transport		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		Covariance							
Construction	Lagged variance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Transport		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Construction	Lagged covariance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Transport	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Table A6: Fraction of regressors (rows) selected to equations (column) to adaptive LassoVAR(1) based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.35	0.45	0.14	0.10	0.33	0.04	0.18	0.35
Consumer, Cyclical		0.25	0.31	0.22	0.15	0.15	0.10	0.31	0.26
Energy		0.31	0.35	0.25	0.09	0.22	0.15	0.10	0.26
Financial		0.25	0.23	0.31	0.37	0.18	0.14	0.11	0.20
Industry		0.21	0.21	0.22	0.17	0.20	0.29	0.18	0.26
Oil & Mining		0.27	0.10	0.27	0.13	0.24	0.29	0.15	0.06
Services		0.10	0.08	0.22	0.27	0.23	0.24	0.35	0.24
Transport		0.21	0.14	0.19	0.14	0.26	0.29	0.28	0.15
Construction	Lagged covariance	0.07	0.05	0.06	0.05	0.05	0.03	0.07	0.03
Consumer, Cyclical		0.07	0.06	0.05	0.07	0.06	0.06	0.06	0.06
Energy		0.09	0.10	0.09	0.10	0.08	0.08	0.09	0.09
Financial		0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03
Industry		0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06
Oil & Mining		0.07	0.06	0.06	0.07	0.06	0.04	0.06	0.06
Services		0.03	0.04	0.06	0.05	0.05	0.05	0.05	0.05
Transport		0.06	0.05	0.06	0.06	0.06	0.05	0.04	0.06
		Covariance							
Construction	Lagged variance	0.17	0.14	0.10	0.10	0.16	0.13	0.10	0.11
Consumer, Cyclical		0.21	0.23	0.23	0.20	0.19	0.21	0.23	0.25
Energy		0.22	0.13	0.18	0.13	0.15	0.14	0.14	0.11
Financial		0.27	0.24	0.29	0.31	0.15	0.18	0.23	0.21
Industry		0.18	0.17	0.19	0.19	0.16	0.17	0.17	0.19
Oil & Mining		0.15	0.13	0.12	0.12	0.16	0.13	0.14	0.12
Services		0.26	0.25	0.20	0.32	0.20	0.21	0.25	0.28
Transport		0.21	0.18	0.13	0.19	0.22	0.22	0.24	0.16
Construction	Lagged covariance	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05
Consumer, Cyclical		0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
Energy		0.08	0.08	0.09	0.08	0.08	0.08	0.08	0.09
Financial		0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Industry		0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Oil & Mining		0.06	0.06	0.07	0.06	0.06	0.06	0.06	0.06
Services		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Source: The authors.

Transport	0.06	0.06	0.07	0.06	0.06	0.06	0.07	0.07

Source: The authors.

Table A7: Fraction of regressors (rows) selected to equations (column) to Lasso-VAR(5) based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
					Variance				
Construction	Lagged variance	0.07	0.09	0.09	0.08	0.06	0.06	0.09	0.11
Consumer, Cyclical		0.07	0.07	0.03	0.06	0.06	0.05	0.07	0.08
Energy		0.11	0.10	0.10	0.05	0.04	0.04	0.05	0.04
Financial		0.08	0.03	0.03	0.07	0.02	0.03	0.06	0.03
Industry		0.05	0.04	0.05	0.06	0.04	0.06	0.06	0.05
Oil & Mining		0.02	0.02	0.02	0.03	0.05	0.05	0.04	0.02
Services		0.02	0.03	0.06	0.03	0.04	0.02	0.05	0.02
Transport		0.07	0.05	0.11	0.06	0.07	0.09	0.07	0.06
Construction	Lagged covariance	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.02
Financial		0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Transport		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
					Covariance				
Construction	Lagged variance	0.08	0.07	0.10	0.09	0.05	0.07	0.07	0.09
Consumer, Cyclical		0.08	0.07	0.06	0.06	0.07	0.06	0.07	0.06
Energy		0.07	0.04	0.03	0.04	0.05	0.03	0.03	0.04
Financial		0.05	0.04	0.04	0.05	0.03	0.04	0.04	0.04
Industry		0.04	0.04	0.04	0.05	0.04	0.05	0.04	0.04
Oil & Mining		0.04	0.05	0.03	0.05	0.04	0.03	0.04	0.04
Services		0.04	0.03	0.03	0.04	0.03	0.02	0.03	0.03
Transport		0.07	0.07	0.06	0.07	0.08	0.08	0.09	0.07
Construction	Lagged covariance	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Consumer, Cyclical		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Energy		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Financial		0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00
Industry		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Oil & Mining		0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Services		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Transport	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Source: The authors.

Table A8: Fraction of regressors (rows) selected to equations (column) to adaptive LassoVAR(5) based models

Sectors		Construction	Consumer, Cyclical	Energy	Financial	Industry	Oil & Mining	Services	Transport
		Variance							
Construction	Lagged variance	0.12	0.13	0.14	0.13	0.12	0.12	0.15	0.19
Consumer, Cyclical		0.13	0.13	0.08	0.12	0.12	0.11	0.12	0.14
Energy		0.19	0.16	0.15	0.08	0.09	0.08	0.08	0.10
Financial		0.14	0.07	0.06	0.12	0.04	0.05	0.10	0.05
Industry		0.09	0.08	0.10	0.10	0.09	0.10	0.10	0.10
Oil & Mining		0.04	0.06	0.04	0.07	0.08	0.09	0.07	0.04
Services		0.05	0.07	0.10	0.08	0.07	0.05	0.09	0.04
Transport		0.13	0.10	0.16	0.13	0.11	0.14	0.12	0.11
Construction	Lagged covariance	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03
Consumer, Cyclical		0.02	0.02	0.03	0.02	0.03	0.02	0.02	0.02
Energy		0.04	0.04	0.03	0.04	0.04	0.04	0.04	0.04
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Oil & Mining		0.02	0.03	0.02	0.03	0.02	0.02	0.03	0.03
Services		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Transport		0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02
		Covariance							
Construction	Lagged variance	0.13	0.11	0.16	0.14	0.09	0.13	0.12	0.15
Consumer, Cyclical		0.14	0.13	0.10	0.12	0.12	0.12	0.13	0.11
Energy		0.12	0.09	0.06	0.07	0.10	0.06	0.07	0.08
Financial		0.09	0.07	0.07	0.09	0.06	0.07	0.07	0.07
Industry		0.09	0.08	0.09	0.09	0.08	0.09	0.08	0.09
Oil & Mining		0.07	0.08	0.06	0.09	0.07	0.06	0.08	0.07
Services		0.08	0.07	0.07	0.07	0.07	0.06	0.06	0.06
Transport		0.12	0.12	0.11	0.13	0.13	0.14	0.14	0.12
Construction	Lagged covariance	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
Consumer, Cyclical		0.03	0.03	0.03	0.02	0.03	0.02	0.03	0.03
Energy		0.03	0.03	0.04	0.04	0.03	0.04	0.03	0.04
Financial		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Industry		0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.03
Oil & Mining		0.02	0.02	0.03	0.03	0.02	0.03	0.03	0.03
Services		0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Source: The authors.

Transport	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Source: The authors.

4.6 Apresentações do Seminário

- Apresentação - Marcelo Kfoury (EESP-FGV): slides
- Apresentação - Luis Fernando Azevedo (MZK): slides
- Apresentação - Patrícia Palomo (Sonata Gestora de Recursos): slides
- Apresentação - Sergio Castelani e Danilo Igliori (Grupo Zap) : slides
- Apresentação - Emerson Marçal (CEMAP-EESP-FGV) : slides
- Apresentação - Pedro Valls (CEQEF-EESP-FGV) : slides