

Forecasting Industrial Production Index by its aggregated or disaggregated data? Evidence from one important emerging market economy*

Diogo de Prince[†]

Emerson Fernandes Marçal[‡]

Pedro Luiz Valls Pereira[§]

Abstract

Our work aims to address if the use of disaggregate data helps to forecasting industrial production index. We use Brazilian industrial production data and we investigate if disaggregate information improves the accuracy of the forecasts. We use a number of recent econometric techniques such as the weighted lag adaptative least absolute shrinkage and selection operator (WLadaLASSO) methodology, the exponential smoothing (selecting the most appropriate model) and Autometrics algorithm to model both aggregates and disaggregates. As far as we know this is the novelty of the work. We run a forecasting exercise from one up to 12 months ahead for Brazilian industrial production. Our full sample covers the period from January of 2002 to August of 2017. Our results suggest that modeling disaggregate data better using exponential smoothing model provides the best performance for 1 up to 7 months ahead for Brazilian industrial production using mean square error as a metric and Autometrics algorithm provides better forecast for 8 up to 12 months but it is not clear whether aggregate or disaggregate data is the best choice given that they are both part of the final set of good predictions.

JEL Codes: C53, E27, C52

Key Words: industrial production, forecasting, model selection.

1 Introduction

Economic agents decide based on a global view of how the economy behaves at that moment and what are your expectations for the future. The general levels of output, employment, interest rates, exchange rates and inflation are examples of key economic indicators that help in the country's diagnosis. Therefore, the proposition and evaluation of econometric models devoted to forecast bring benefits to build better guides to economic agents and policy makers.

One of the most comprehensive and important macroeconomic indicators of the economy is the Gross Domestic Product (GDP), as it is a proxy for a country's economic performance. In the present work, the proxy used for GDP is industrial production, since the indicator of industrial production is monthly (higher frequency than GDP) and is released with a lag of about one month, which is therefore smaller than GDP (with a delay of more than two months).

*We thank seminar participants at Sao Paulo School of Economics and Anpec 2018 (Rio de Janeiro). The authors gratefully

acknowledge financial support from CNPq. We also thank Eusebio Souza and Thais Beserra for the research assistance.

[†] Researcher of Center for Applied Macroeconomic Research and Assistant Professor at Federal University of Sao Paulo

[‡] Lecturer at Sao Paulo School of Economics and Head of Center for Applied Macroeconomic Research

[§]Full Professor at Sao Paulo School of Economics and Head of Center for Research in Quantitative Finance and Economics

A point to address is if the use of disaggregated data improves to forecast variables comparable to using aggregated series. Data disaggregation is an alternative to lead to more accurate forecasts. This alternative refers to the decomposition of the main variable into several subcomponents, which have different weights for the aggregate series. We estimate these subcomponents individually and then we grouped these subcomponents to obtain a forecast of the aggregate series. This technique can increase the quality of the forecast because we model the subcomponents taking into account their individual characteristics. We use this alternative in the present work to understand if there are gains to predict the aggregate series if we estimate each subcomponent and we use the weight of this subcomponent. Another alternative is to estimate using only lagged aggregate variable to forecast the aggregate series.

The accuracy of using disaggregated or aggregate data for forecasting has been discussed in some studies. The theoretical studies indicate that when the data generating process (DGP) is known, it is preferable to first use the disaggregated data in multiple series to later aggregate them than to directly predict the already aggregated series. However, the literature acknowledges that in most cases DGP is unknown and therefore the use of aggregate series may be preferable due to variability of specification and estimation of the model (because the estimation has fewer parameters). Some examples of contributions to the theoretical literature on aggregate or disaggregate forecasting are Lutkepohl (1984, 1987), Granger (1987), Pesaran, Pierse, and Kumar (1989), Garderen, Lee, and Pesaran (2000), and Giacomini and Granger (2004). As DGP is not known, then the question (if aggregating the disaggregated forecasts improves the accuracy of the aggregate forecast) becomes an empirical question.

Our point is whether the prediction of the disaggregated components of Brazilian industrial production improves the accuracy of the forecast of aggregate Brazilian industrial production. Our contribution is that we do not know articles that address the contribution of the disaggregated data of the weighted lag adaptive least absolute shrinkage and selection operator (WLadaLASSO) methodology or the exponential smoothing (selecting the most appropriate model). In addition, there are few papers that analyze the contribution of disaggregated data to forecast the industrial production and we intend to fill this gap. Using monthly data from 2002 to 2017, we select the best univariate model, we estimate a rolling window of 100 fixed observations and we evaluate the forecast from 1 to 12 months ahead for Brazilian industrial production, in which we estimate 61 rolling windows. We consider as naive models the first order autoregressive model (AR(1)), AR(1) with time-varying parameters (TVP AR(1)), and the Stock and Watson (2007) unobserved components with stochastic volatility (UC-SV) estimated based on Barnett et al (2014). We consider the following methods for selecting the best model: exponential smoothing based on Hyndman et al (2002), Hyndman and Khandakar (2008) and Hyndman et al (2008), the least absolute shrinkage and selection operator (LASSO), adaptive LASSO (adaLASSO), the WLadaLASSO and the Autometrics algorithm with break dummy variables. We use the LASSO and its variants to select the lags from an AR (15) and we consider the Autometrics algorithm to select break dummy variables in the sample and the lags from an AR(15). We compare the prediction performance between the models based on the mean square error (MSE), the Diebold and Mariano (1995) test and the Model Confidence set procedure of Hansen et al (2011).

Our results point to a better performance of the exponential smoothing model with disaggregated data for the forecast of 1 to 7 months ahead for Brazilian industrial production by MSE. But the Autometrics algorithm with step dummy variables (and considering a p-value of 0.1%) with the disaggregated forecasts has more accuracy for 10 to 12 months ahead by MSE. To analyze whether there is a better statistical performance, we use the exponential model with disaggregated data as a benchmark in the Diebold and Mariano (1995) test. We obtain that the disaggregated exponential smoothing model presents better forecast performance in relation to the naive models (AR(1), AR(1) with time-varying parameters, UC-SV) considered with aggregate or disaggregated data and the aggregate exponential smoothing model.

The article structure contains six sections in addition to this introduction. The following is a brief review of the literature. Next, we address the methodology of the models considered in the paper. Section 4 presents the

data used, the empirical forecasting strategy, the Diebold and Mariano (1995) test and the Model Confidence set to compare the performance of the models. In section 5, we discuss the results of the study. Finally, we show the final considerations of the research and the next steps.

2 Literature Review

This section presents some empirical articles that address the difference in forecast accuracy between aggregating the disaggregated forecasts or modeling only the aggregate variable. Stock and Watson (1998) compare 49 univariate projection models to forecast industrial activity and inflation in the United States from 215 monthly series from the years 1959 to 1996. One conclusion of the paper is that the forecast aggregation has better performance than separate forecasts. The authors also find significant gains using forecast combinations and these gains are enough to justify their use by a risk-averse analyst. Stock and Watson (1998) also point to the importance of performing the unit root test to reduce errors substantially in the estimates. Tobias and Zellner (1998) seek to determine the effects of aggregation and disaggregation to predict the average annual growth rate of 18 countries. In general, the disaggregation leads to more observations to estimate the parameters, besides the authors obtain better predictions for the aggregate variable (growth rate).

Marcellino, Stock and Watson (2003) find evidence that the individual estimation of inflation in each euro area country and the subsequent aggregation of projections increases the accuracy of the final result of the estimation in relation to the option to forecast this variable only in aggregate level. Hubrich (2005) obtains that aggregating forecasts for each component of inflation does not necessarily better predict year-on-year inflation in the euro area one year ahead. Espasa, Senra and Albacete (2002) have similar results indicating, however, that the disaggregation leads to better projections for periods longer than one month. Carlos and Marial (2016) compare forecasts from models for aggregate inflation and aggregating the forecasts for the groups and items of the Brazilian inflation index. The authors obtain that there are gains in the accuracy of the forecast with disaggregated data.

Barhoumi et al (2010) analyze the forecasting performance of the France GDP between alternative factor models. The point of the work is whether it is more appropriate to extract factors from aggregate or disaggregated data to forecast. Rather than using 140 disaggregated series, Barhoumi et al (2010) show that the static approach of Stock and Watson (2002) with 20 aggregate series leads to better prediction results. The next section presents the methodology we use in this paper.

Regarding the literature on the methodologies used in this work, Epprechat et al (2019) conduct a Monte Carlo simulation experiment considering the DGP as a linear regression with orthogonal variables and independent data. The DGP with orthogonal variables favors LASSO-type models for the Monte Carlo simulation exercise because a LASSO assumption is the irrepresentable condition (Zhao, Yu, 2006).¹ According to this condition, the relevant variable can not be very correlated with the irrelevant variable. The authors obtain a result that adaLASSO and Autometrics have similar forecasting performance with small values of relevant variables and when the candidate variables are lower than the number of observations. The Autometrics algorithm only performs better when they have a large number of relevant variables because of the bias from the penalization term in adaLASSO. Also Epprechat et al (2019) obtain that adaLASSO has superior performance in model selection than LASSO and Autometrics for a linear regression with orthogonal variables. Only when we have small samples that it is preferable to use the Autometrics. The authors also have a genomic data to compare the predictive power to epidermal thickness in psoriatic patients, in which covariates are not orthogonal. Out-of-sample forecasts with the selection of variables by LASSO, adaLASSO or Autometrics can not be statistically differentiated by the modified Diebold and Mariano (1995) test, proposed by Harvey et al (1997).

Kock and Teresvirta (2014) consider the neural network model with three algorithms to model monthly

¹AdaLASSO requires a weighted irrepresentable condition, which is a less restrictive condition than irrepresentable condition.

industrial production and unemployment series from the Group of Seven (G7) countries and Denmark, Finland, Norway and Sweden. They focus on forecasting during the economic crisis 2007-2009. The authors find that Autometrics algorithm performs worse with direct forecast than with recursive forecast because the model is not a reasonable approximation of reality by excluding the most relevant lags. Autometrics algorithm tends to select a highly parameterized model that does not present competitive forecasts compared to other methodologies in the direct forecast. That is, Kock and Teräsvirta (2014) obtain that the Autometrics algorithm may perform worse under considerable misspecification of the general model.

3 Methodology

We consider three naive estimators to compare our model forecasts: UC with stochastic volatility, autoregressive model and time-varying parameters autoregressive model. Our point is select the lags of the variable that are relevant based on different univariate methodologies. We use LASSO and two variants (adaLASSO and WLadaLASSO), exponential smoothing method, and Autometrics algorithm. We use the variable y_t in the case of the AR (1), TVP-AR (1), LASSO and its variants, and the Autometrics algorithm because the series y_t is non-stationary. In the case of UC with stochastic volatility and exponential smoothing we can use the variable y_t .

3.1 Time-varying parameters autoregressive model of first order

In this section, we present the methodology of the first order of the autoregressive model with time-varying parameters. We use this model to obtain a naive forecasting and we consider only one lag in the model. The methodology with time-varying parameters seeks to contemplate the changes that can occur in the economy over time (Kapetanios et al, 2017). Consider the time-varying parameter AR(1) model, where we can write the measurement equation as

$$y_t = \beta_{0t} + \beta_{1t} y_{t-1} + \varepsilon_t \quad (1)$$

where $\varepsilon_t \sim N(0, \sigma^2)$ for $t = 1, \dots, T$, and y_0 is an initial observation. We can write the autoregressive and the constant coefficients $\beta_t = (\beta_{0t}, \beta_{1t})$ with the following transition equation

$$\beta_t = \beta_{t-1} + u_t \quad (2)$$

where $u_t \sim N(0, \Omega)$ and the transition equation has the initial value $\beta_1 \sim N(\beta_0, \Omega_0)$.

We use Bayesian estimation following Kroese and Chan (2014) and for this we stack the observations over all times t from the matrix notation of the equation (1). So, we have

$$y = X\beta + \varepsilon \quad (3)$$

where $y = (y_1, \dots, y_T)$, $\beta = (\beta_1, \dots, \beta_T)$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T) \sim N(0, \sigma^2 I)$, $X = \begin{matrix} & & & x_t & 0 & \dots & 0 \\ & & & 0 & x_t & \dots & 0 \\ & & & \vdots & \vdots & \ddots & \vdots \\ & & & \vdots & \vdots & \ddots & \vdots \\ & & & 0 & 0 & \dots & x_t \end{matrix}$,

and $x_t = (1, y_{t-1})$. The logarithm of the joint density function of y (omitting the initial observation y_0) is

$$\ln f(y|\beta, \sigma^2) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) + \text{const} \quad (4)$$

where const is the constant term. Now we can stack the transition equation (2) over time t . We consider $\beta_0 = 0$

for simplification. We can write the transition equation (2) in matrix form as

$$H\beta = u \quad (5)$$

$$\text{where } u \sim N(0, S), u = (u_1, \dots, u_T), H = \begin{pmatrix} I & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \text{ and } S = \begin{pmatrix} \Omega_0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \Omega & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \Omega \end{pmatrix}.$$

Given that $|H| = 1$ and $|S| = |\Omega_0| |\Omega|^{T-1}$, the logarithm of the joint density function of β is given by

$$\ln f(\beta | \Omega) = -\frac{T-1}{2} \ln |\Omega| - \frac{1}{2} \beta^T H S^{-1} H \beta + \text{const} \quad (6)$$

We can reduce the number of parameters assuming that Ω is diagonal. So consider $\omega^2 = (\omega_1^2, \omega_2^2, \dots, \omega_p^2)$ as the vector of diagonal elements of Ω .

We can obtain the posterior density specifying the prior for β^2 and σ^2 . Assume an independent prior $f(\sigma^2) f(\omega^2)$, where $\sigma^2 \sim IG(\alpha_{\sigma^2}, \lambda_{\sigma^2})$, $\omega_i^2 \sim IG(\alpha_{\omega_i^2}, \lambda_{\omega_i^2})$ and IG is the inverse-gamma distribution. We specify the constants α_{σ^2} , λ_{σ^2} , $\alpha_{\omega_i^2}$ and $\lambda_{\omega_i^2}$.

The posterior density function is given by

$$f(\beta, \sigma^2, \omega^2 | y) \propto f(y | \beta, \sigma^2) f(\beta | \omega^2) f(\sigma^2) f(\omega^2) \quad (7)$$

where $\ln f(\beta | \sigma^2)$ and $\ln f(\beta | \omega^2)$ are given by (4) and (6) respectively. We can obtain posterior draws using Gibbs sampler. We draw from $f(\beta | y, \sigma^2, \omega^2)$ followed by a draw from $f(\sigma^2, \omega^2 | y, \beta)$. As $f(\beta | y, \sigma^2, \omega^2)$ is a normal density, if we determine the mean vector and the precision matrix, we can apply the algorithm described below to obtain a draw from it efficiently. Using (4) and (6), we write

$$\ln f(\beta | y, \sigma^2, \omega^2) = \ln f(y | \beta, \sigma^2) + \ln f(\beta | \omega^2) + \text{const} \quad (8)$$

as

$$\ln f(\beta | y, \sigma^2, \omega^2) = -\frac{1}{2} (\beta - \hat{\beta})^T K_\beta (\beta - \hat{\beta}) + \text{const} \quad (9)$$

where $K_\beta = \frac{1}{\sigma^2} X^T X + H S^{-1} H$ and $\hat{\beta} = K_\beta^{-1} (\frac{1}{\sigma^2} X^T y)$. This means that $\beta | y, \sigma^2, \omega^2 \sim N(\hat{\beta}, K_\beta^{-1})$. Then with the algorithm described soon we can draw $\beta | y, \sigma^2, \omega^2$.

The algorithm generates the multivariate normal vector generation using the precision matrix. The algorithm obtains N independent draws from $N(\mu, \Lambda^{-1})$ of dimensions n with the following steps:

1. We obtain the lower Cholesky factorization $\Lambda = DD^T$.
2. We draw $Z_1, \dots, Z_n \sim N(0, 1)$.
3. We determine Y from $Z = D^T Y$.
4. We obtain $W = \mu + Y$.
5. We repeat steps 2-4 independently N times.

The next point is to be able to draw $(\sigma^2, \omega^2 | y, \beta)$. Given y and β , σ^2 and ω^2 are conditionally independent. From (7), we have $f(\sigma^2 | y, \beta) \propto f(y | \beta, \sigma^2) f(\sigma^2)$ and $f(\omega^2 | y, \beta) \propto f(\beta | \omega^2) f(\omega^2)$. Both conditional densities are inverse-gamma densities, which shows that

$$\sigma^2 | y, \beta \sim IG\left(\alpha_{\sigma^2} + \frac{T}{2}, \lambda_{\sigma^2} + \frac{1}{2} (y - X\beta)^T (y - X\beta)\right) \quad (10)$$

and

$$\omega_i | y, \beta \sim IG \left(\alpha_{\omega^2} + \frac{T-1}{2}, \frac{1}{2} \sum_{t=2}^T (\beta_{it} - \beta_{it-1})^2 \right) \quad (11)$$

Following Kroese and Chan (2014), we set small values for the shape parameter of the inverse gamma distribution so that the prior is more non-informative. That is, $\alpha_{\omega^2} = \alpha_{\omega^2} = 5$, $i = 0, 1$. Also we set the prior as $\lambda_{\omega^2} = (\alpha_{\omega^2} - 1)$, $\lambda_{\omega^2} = 0.1^2 \alpha_{\omega^2} - 1$, and $\lambda_{\omega^2} = 0.1^2 \alpha_{\omega^2} - 1$. Finally, we set the covariance matrix Ω_0 to be diagonal with diagonal elements equal to five, in line with Kroese and Chan (2014).

3.2 UC with stochastic volatility

Stock and Watson (2007) include stochastic volatility in an unobserved component model. The authors show that UC-SV presents a great performance to forecast US inflation. The UC-SV model is defined as

$$y_t = \beta_t + \sigma_t^{\frac{1}{2}} v_t \quad (12)$$

$$\beta_t = \beta_{t-1} + \omega_t^{\frac{1}{2}} e_t \quad (13)$$

where $\ln \sigma_t$ and $\ln \omega_t$ are the logarithm of the stochastic volatility, β_t is the trend, $\ln \sigma_t = \ln \sigma_{t-1} + e_{1t}$, and $\ln \omega_t = \ln \omega_{t-1} + e_{2t}$, in which the variances of e_{1t} and e_{2t} are respectively g_1 and g_2 . We estimate the model using Markov chain Monte Carlo (MCMC) algorithm with Gibbs sampling method following Barnett et al (2014).

Our first step is to establish the priors and starting values. We define the prior for the initial value of the $\ln \sigma_t$ as $\ln \sigma_0 \sim N(\mu_0, 10)$ where μ_0 is the variance of $y_{t_0} - \beta_{t_0}$ and t_0 refers to the training sample of 40 observations and β_{t_0} is an initial estimate for the trend from the Hodrick-Prescott filter. In a similar way, $\ln \omega_0 \sim N(\omega_0, 10)$ where $\omega_0 = \Delta \beta_{t_0}$. We use the priors for g_1 and g_2 from an inverse gamma, we set the prior scale parameter equal to 0.01 and 0.0001 respectively with one degree of freedom as Barnett et al (2014). So, we use non-informative priors.

So the next step is to simulate the posterior distributions. We draw σ_t and ω_t conditional on the value for g_1 and g_2 with the Metropolis algorithm based on Jacquier et al (2004). We draw β_t using the Carter and Kohn (2004) algorithm. We generate the sample for g_1 and g_2 from the inverse gamma distribution. We consider 10,000 replications of the MCMC algorithm and we keep the last 1,000 replications for inference.

Below we detail how we calculate the marginal likelihood. We use a particle filter to calculate the log likelihood function for the UC-SV. We define Ξ as all parameters of the model. Based on Chib (1995), we consider the log marginal likelihood as:

$$\ln P(y_t) = \ln F(y_t | \hat{\Xi}) + \ln P(\hat{\Xi}) - \ln G(\hat{\Xi} | y_t) \quad (14)$$

where $\ln P(y_t)$ is the log marginal likelihood that we want to calculate, $\ln F(y_t | \hat{\Xi})$ is the log likelihood function, $\ln P(\hat{\Xi})$ is the log prior density, and $\ln G(\hat{\Xi} | y_t)$ is the log posterior density of the model parameters. The three elements on the right hand side of (14) are evaluated at the posterior mean for the model parameters $\hat{\Xi}$.

We calculate the log likelihood function for this model using a particle filter following Barnett et al (2014). But we need an additional step to obtain the term $\ln G(\hat{\Xi} | y_t)$. $\ln G(\hat{\Xi} | y_t)$ can be factorized into conditional and marginal densities of various parameter blocks and we use Gibbs and Metropolis algorithm to approximate these densities according to Chib (1995) and Chib and Jeliazkov (2001). The posterior density is defined as $G(\hat{\Xi} | y_t) = G(\hat{g}_1, \hat{g}_2)$ and we drop the dependence on y_t to simplify the notation. The factorization of this density can be described by

$$G(\hat{g}_1, \hat{g}_2) = H(\hat{g}_1 | \hat{g}_2) H(\hat{g}_2) \quad (15)$$

where $H(\hat{g}_1|\hat{g}_2) = \int H(\hat{g}_1|\hat{g}_2, \Theta) H(\Theta|\hat{g}_2) d\Theta$ and $H(\hat{g}_2) = \int H(\hat{g}_2|\Theta) H(\Theta) d\Theta$, in which $\Theta = \{\beta_t, \sigma_t, \omega_t\}$ is the state variables in the model. These two densities $H(\hat{g}_1|\hat{g}_2)$ and $H(\hat{g}_2)$ can be obtained as a 'weighted average' across state variables. We can approximate $H(\hat{g}_1|\hat{g}_2)$ and $H(\hat{g}_2)$ with additional Gibbs runs and we can integrate over the states. We consider 10,000 iterations in these additional Gibbs samplers and we remain with the last 3,000 like Barnett et al (2014).

3.3 Lasso-type penalties

We present three lasso-type penalties in this subsection to select the relevant lags of the univariate model from an AR (15): LASSO, adaLASSO, and WLadaLASSO.

3.3.1 Lasso

Tibshirani (1996) proposes the LASSO method based on the following minimization problem

$$\hat{\beta}^{LASSO} = \underset{\beta_0, \beta_1, \dots, \beta_k}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ji})^2 + \lambda \sum_{j=1}^k |\beta_j| \quad (16)$$

where $\lambda \geq 0$ is a tuning parameter and LASSO requires a method to obtain a value for λ , that we explain soon.

The first term is the sum of square of residuals and the second term is a shrinkage penalty. $|\beta_j|$ is the ℓ^1 norm of a coefficient vector β . The ℓ^1 penalty forces some of the coefficients estimates to be equal to zero when λ is sufficiently large. When $\lambda = 0$, LASSO estimates are equal to ordinary least squares estimates. So, LASSO technique performs variable selection.

Cross-validation is usually the method to obtain the λ value. With time series data, we use the Bayesian information criterion (BIC) to choose λ , following Konzen and Ziegelmann (2016). We consider a grid of λ values to choose one value.

3.3.2 AdaLASSO

Zou (2006) states that LASSO can lead to inconsistent selection of variables that keep noisy variables for a given λ that leads to optimal estimation rate. Also the author shows that LASSO can lead to the right selection of variables with biased estimates for large coefficients and this take to suboptimal prediction rates.

So, Zou (2006) introduces the adaptive LASSO, which considers weights ω_j that adjust the penalty to be different for each coefficient. The adaptive LASSO seeks to minimize

$$\hat{\beta}^{adaLASSO} = \underset{\beta_0, \beta_1, \dots, \beta_k}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ji})^2 + \lambda \sum_{j=1}^k \omega_j |\beta_j| \quad (17)$$

where $\omega_j = |\hat{\beta}_j^{ridge}|^{-\tau}$, $\tau > 0$. The adaptive LASSO considers that large (small) coefficients have small (large) weights - and small (large) penalties. The coefficients estimated by ridge regression $\hat{\beta}^{ridge}$ lead to get the weight ω_j . Ridge regression shrinks the vector of coefficients by penalizing the sum of the squares of the residuals:

$$\hat{\beta}^{ridge} = \underset{\beta_0, \beta_1, \dots, \beta_k}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ji})^2 + \lambda \sum_{j=1}^k \beta_j^2 \quad (18)$$

where the penalty is the ℓ^2 norm of the β vector. Ridge regression is not a method of variable selection because

this regression obtains non-zero estimates for all coefficients.

3.3.3 WLadallasso

When we use adaLASSO with time series data, each coefficient associated with a lagged variable is penalized according to the size of the ridge's estimate. The less distant the lag of the variable, the more important the variable must be for the model and therefore its coefficient should be less penalized (considering the case without seasonality).

Park and Sakaori (2013) propose some types of penalties for different lags. Konzen and Ziegelmann (2016) present the adaLASSO with weighted lags based on Park and Sakaori (2013), denominated as WLadaLASSO. The WLadaLASSO method is given by

$$\hat{\beta}^{wladaLASSO} = \underset{\beta_0, \beta_1, \dots, \beta_k}{\operatorname{argmin}} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ji})^2 + \lambda \sum_{j=1}^k \omega_j |\beta_j| \quad (19)$$

where $\omega_j = |\hat{\beta}_j^{ridge}| e^{-\alpha l} e^{-\tau}$ is the weight, $\tau > 0$, $\alpha \geq 0$, and l is the order of the variable's lag. We set the parameter τ equal to one as Konzen and Ziegelmann (2016) for the adaLASSO and WLadaLASSO. We consider a grid for α , where the set of possible values for α is $[0, 0.5, 1, \dots, 10]$. We calculate the optimal λ among those possible for that model with the lowest BIC value for each value of α . We choose the α value from that model that produces the smallest BIC value among all possible α values, following Konzen and Ziegelmann (2016).

3.4 Exponential smoothing

The name exponential smoothing comes from the weights decreasing exponentially when the observation becomes older. Exponential smoothing basically is an exponentially weighted sums of past values to obtain the forecast. We can represent the exponential smoothing methods as state space models (Ord et al, 1997, Hyndman et al, 2002, Hyndman et al, 2005). The exponential smoothing method is an algorithm which only produces point forecasts. The stochastic state space model associated to this method provides a framework which leads to the same point forecast but also estimates the prediction intervals for example (Hyndman et al, 2008).

Economic series exhibit some features that we can work on. We can decompose the economic series into certain components, such as trend (T), cycle (C), seasonality (S), and irregular or error (E). We use the exponential smoothing method to decompose into these components. Only the cycle component is not decomposed separately so that we model along with the trend component, following Hyndman et al (2005), Hyndman and Khandakar (2008), and Hyndman et al (2008). Thus, we can combine the components of trend, seasonality and error.

Hyndman et al (2002), Hyndman and Khandakar (2008) and Hyndman et al (2008) propose 15 different combinations between trend and seasonality components. The trend component can present five different possibilities: none, additive, additive damped, multiplicative, and multiplicative damped. The trend component is a combination between the level (l) and growth (b) parameter. Consider T_h the forecast trend over the next h periods, and φ is a damping parameter ($0 < \varphi < 1$). So if there is no trend component, $T_h = l$. If the trend component is additive, $T_h = l + bh$. If the trend component is additive damped, $T_h = l + (\varphi + \varphi^2 + \dots + \varphi^h) b$. If the trend component is multiplicative, $T_h = l b^h$. If the trend component is multiplicative damped, $T_h = l b^h (\varphi + \varphi^2 + \dots + \varphi^h)$. If growth rate at the end of the series is unlikely to continue, the damped trend seems to be a reasonable option.

After we present the types of trend component, the next step is to detail the types of seasonal component. The seasonal component can be none, additive or multiplicative. Also the error component can be additive or multiplicative, but this distinction is not relevant to make point forecast (Hyndman et al, 2008).

So, we consider the combination of five types of trend component and three types of seasonal component which leads to a total of 15 types of exponential smoothing methods that we consider in this paper, following Hyndman and Khandakar (2008). We present these 15 possibilities in the Table 1, in which the first entry refers to the trend

component and the second to the seasonal component. Some of these exponential smoothing methods are known by other names. For example, cell N , N represents the simple exponential smoothing method, cell A, N refers to the Holt's linear method, and cell A_d, N is associated with the damped trend method. The cell A, A describes the additive Holt-Winter's method, the cell A, M refers to the multiplicative Holt-Winter's method.

Trend component	Seasonal component		
	N (None)	A (Additive)	M (Multiplicative)
N (None)	N,N	N,A	N,M
A (Additive)	A,N	A,A	A,M
A_d (Additive damped)	A_d,N	A_d,A	A_d,M
M (Multiplicative)	M,N	M,A	M,M
M_d (Multiplicative damped)	M_d,N	M_d,A	M_d,M

Table 1: The different combinations of exponential smoothing methods

For example, consider the Holt's linear method (cell A, N) that can be described as

$$\hat{y}_t = \alpha y_t + (1 - \alpha)(\hat{y}_{t-1} + b_{t-1}) \quad (20)$$

$$b_t = \beta^*(\hat{y}_t - \hat{y}_{t-1}) + (1 - \beta^*)b_{t-1} \quad (21)$$

$$\hat{y}_{t+h|t} = \hat{y}_t + b_t h \quad (22)$$

where equation (20) shows the model to the level of the series at time t \hat{y}_t in this case. The equation (21) describes the growth rate (slope estimate) of the series at time t b_t . b_t is a weighted average of the estimate of growth obtained by the difference between successive levels and the previous growth b_{t-1} . Finally, the equation (22) presents the prediction for the variable y h periods ahead using information available at time t . This equation describes that the forecast for the variable h periods ahead is given by the level in the current time \hat{y}_t adding the growth b_t for the h periods. α is the smoothing parameter for the level with $0 < \alpha < 1$, and β^* is the smoothing parameter for the trend with $0 < \beta^* < 1$.

Now we consider a model with the seasonality component. When seasonal variations are constant throughout the series, the additive method for the seasonal component is preferred. When seasonal variations change proportionally to the level of the series, the multiplicative method for the seasonal component is preferred. Consider for example the Holt-Winters method of additive trend with additive seasonal component (cell A, A), the equations of this method are given by

$$\hat{y}_t = \alpha (y_t - s_{t-m}) + (1 - \alpha)(\hat{y}_{t-1} + b_{t-1}) \quad (23)$$

$$b_t = \beta^*(\hat{y}_t - \hat{y}_{t-1}) + (1 - \beta^*)b_{t-1} \quad (24)$$

$$s_t = \gamma (y_t - \hat{y}_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (25)$$

$$\hat{y}_{t+h|t} = \hat{y}_t + b_t h + s_{t-m+h}^* \quad (26)$$

The equations (23), (24), (25), and (26) describe, respectively, the level \hat{y}_t , the growth rate b_t , the seasonality s_t , and the forecast h periods ahead of the series $\hat{y}_{t+h|t}$. m is the length of seasonality (e.g., number of months or

quarters in a year), and $h_m = [(h - 1) \bmod m] + 1$. The parameters of the Holt-Winters method (α, β^*, γ) are restricted to lie between 0 and 1.

Table 2 presents the equations for the level, growth, seasonality, and forecast of the series for h periods ahead for the 15 cases considered. Some values for exponential smoothing parameters lead to interesting specific cases. Some examples are: the level remains constant over time if $\alpha = 0$, the slope is constant over time if $\beta^* = 0$, and the seasonal behavior is the same over time if $\gamma = 0$. Finally, the methods A and M for the trend component are particular cases of A_d and M_d with $\phi = 1$.

Trend	Seasonal		
	N	A	M
N	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$
	$\hat{y}_{t+h t} = \ell_t$	$\hat{y}_{t+h t} = \ell_t + s_{t-m+h_m^+}$	$\hat{y}_{t+h t} = \ell_t s_{t-m+h_m^+}$
A	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
A _d	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
M	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}$
	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}$
M _d	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}b_{t-1}^\phi$
	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$	$b_t = \beta^*(\ell_t/\ell_{t-1}) + (1 - \beta^*)b_{t-1}^\phi$
	$\hat{y}_{t+h t} = \ell_t b_t^h$	$\hat{y}_{t+h t} = \ell_t b_t^h + s_{t-m+h_m^+}$	$\hat{y}_{t+h t} = \ell_t b_t^h s_{t-m+h_m^+}$
	$\hat{y}_{t+h t} = \ell_t b_t^{\phi h}$	$\hat{y}_{t+h t} = \ell_t b_t^{\phi h} + s_{t-m+h_m^+}$	$\hat{y}_{t+h t} = \ell_t b_t^{\phi h} s_{t-m+h_m^+}$

Table 2: The equations for the level, growth, seasonality, and forecast of the series for h periods ahead for the 15 cases considered

In the sequence we present the state space models that underlie exponential smoothing methods. Each of the 15 models considered consist of a measurement equation that describes the data, and state equations that represent how the unobserved components (level, trend, seasonal) modify over time. The measurement equation together with the state equations are known by state space models.

Hyndman et al (2002) propose to differentiate the behavior of the model with additive errors in relation to the multiplicative errors. If the estimated parameters are the same, the point forecast is not affected if the error is multiplicative or additive, only the prediction interval. When considering the behavior of the error, Hyndman et al (2002) have the triplet (E, T, S) that refers to the three components: error, trend, and seasonality. The model $ETS(A, A, N)$ means that the errors and the trend are additive and that there is no seasonality for example.

If we return to the Holt's linear method presented by equations (20), (21), and (22) with additive errors for example, we have $ETS(A, A, N)$. Assuming that error $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim N(0, \sigma^2)$ and independently distributed, we can write the error correction equations as

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t \quad (27)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t \quad (28)$$

$$b_t = b_{t-1} + \beta \varepsilon_t \quad (29)$$

where $\beta = \alpha\beta^*$. Equation (27) refers to the measurement equation and equations (28) and (29) describe the state equations for the level and growth respectively. Similarly, all 15 exponential smoothing methods presented in the table 2 can be rewritten in the form of state space model with additive or multiplicative errors, see Hyndman et al (2008) for example.

Basically, the estimation procedure is based on estimating the smoothing parameters $\alpha, \beta, \gamma, \phi$ and the initial state variables $/!_0, b_0, s_0, s_{-1}, \dots, s_{-m+1}$ maximizing the likelihood function. The algorithm proposed by Hyndman et al. (2002) also determines which of the 15 ETS models is most appropriate by selecting the model based on the information criterion. The information criterion used to select the most appropriate model are Akaike information criterion (AIC), AIC corrected for small sample bias (AICc), and BIC, according to Hyndman et al (2008).

3.5 Autometrics algorithm

We use the Autometrics algorithm (Doornik and Hendry, 2007, 2009, Doornik, 2009) to address potential instability points and structural changes and we select the lags of the dependent variable that are relevant. The algorithm is based on the approach called the London School of Economics of the general model to the particular. From a general unrestricted model, the algorithm attempts to reduce the unrestricted model through combinations between variables of the general model to evaluate the relevance of the variables to eliminate those irrelevant variables (variables with coefficients that are statistically insignificant). At each step in the attempt to reduce the model, the algorithm performs specification error tests in order to verify the congruence of the models after the eliminations. The purpose of this procedure is to determine the comprehensive and parsimonious model that is a good representation of the local data generating process. According to Hendry and Nielsen (2007), a model is congruent when it satisfies specification error tests for (i) heteroscedasticity, autocorrelation and non-normality, (ii) failure of the weak exogeneity hypothesis, (iii) constant parameters over time. The algorithm allows the number of variables to be greater than the number of observations and it deals with the perfect collinearity generated by the saturation dummy variables that we mention below.

We adopt the Autometrics algorithm with a significance level of 1% and 0.1% using the block method with the inclusion of impulse indicator saturation (IIS) and step indicator saturation (SIS) variables. We include all these variables at each point in time in the regression to analyze whether they are relevant through Autometrics algorithm. That is, if we have a regression with 100 observations in time, the algorithm analyzes the relevance of 100 possibilities for the dummy variable IIS for example. IIS is a dummy variable that is one only at a given point in time and zero otherwise. SIS is a variable equal to one from the beginning of the sample to the specific point in time and zero thereafter. The dummy variable SIS deals with the presence of structural breaks.

The Autometrics algorithm starts from a general model with up to 15 lags of the dependent variable for Brazilian industrial production with the following equation

$$y_t = \rho_i y_{t-i} + (\alpha_l IIS_l + \gamma_l SIS_l) + \theta_s Se_s + \varepsilon_t$$

$\begin{matrix} 15 & & T & & 11 \\ i=1 & & l=1 & & s=1 \end{matrix}$

where SIS_l is a step dummy variable equal to one until $t=l$ and zero otherwise, IIS_l is an impulse variable equal to one only at $t=l$ and zero otherwise. Se_s are the monthly dummy variables for each month s . ε_t is the error term.

4 Data and Empirical Strategy

Our data is the Brazilian industrial production index at general level and your disaggregation by sectors. The data source is the Monthly Industrial Survey of Physical Production (PIM-PF) of the Brazilian Institute of Geogra-

phy and Statistics (IBGE). We use monthly data from January 2002 to August 2017 without seasonal adjustment. We consider the first difference of data to have stationary series, with exception of the UC-SV and ETS models.

We forecast the general industry in a model only with the lags of this series and we compare this with the forecast for each sector and we aggregate according to their weight to obtain the general industry forecast. Thus we use the disaggregated data for the extractive industry and the 25 sectors of the manufacturing industry. However, two sectors printing and reproduction of recordings; and maintenance, repair and installation of machines and equipment only present data from January 2012 and therefore we do not include these sectors in the estimates.² Thus, our disaggregated sample includes the category of extractive industries and 23 sectors of the manufacturing industry. Next we detail the empirical strategy and how we compare the predictions obtained.

4.1 Empirical Strategy and Forecast Comparison

We consider the model up to 15 lags of the dependent variable. Thus, we use a rolling window of 100 fixed observations for each estimation for a forecast horizon of 1 to 12 months ahead. With this window size, we estimate 61 samples for general industry forecast series as for each of the sectors. We reestimate the model using LASSO-type methods, Autometrics algorithm, ETS and naive models for the rolling windows. Thus, we allow a re-specification of the best model according to each methodology for each window. We forecast the production of each sector and we use the weight of each sector to obtain the forecast for the general industry from its components. The forecast window for analysis ranges from September 2011 to August 2017.

We compare our forecasts with the estimates of three naive models. The first is the autoregressive model of the first order, the second is the time-varying parameters autoregressive model of the first order, and the third is the UC with SV. We estimate all models with the data for the general industry (aggregate forecast) and the forecast for the general industry from the disaggregated data (disaggregated forecast). Thus, we have two naive predictions from the AR (1) model (aggregate and disaggregated), two naive predictions from the time varying AR(1) (aggregate and disaggregated), and two naive predictions from UC with SV (aggregate and disaggregated). The naive model serves as reference or benchmark for the other forecasts. The performance analysis for the forecast from the models will be through the mean square error (MSE), the Diebold and Mariano (1995) test and Model Confidence Set of Hansen et al (2011) to determine if there is a model with more accurate prediction for the Brazilian index of general industrial production for the period considered. Next we present the test of Diebold and Mariano (1995).

4.1.1 Diebold and Mariano test

The MSE is the difference between the actual value of the data and the estimated value squared. The MSE is an average of this squared difference for the sample forecasts.

The Diebold and Mariano (1995) test presents if there is any model that has a statistically more accurate prediction for the Brazilian general industrial production for the considered period. Consider a loss function g based on forecasting errors, in which we consider a quadratic loss function. Assuming two models (1 and 2), in which the forecasting errors $\hat{\varepsilon}_{t+h|t}^1$ for h period ahead are given by $\hat{\varepsilon}_{t+h|t}^1 = y_{t+h} - \hat{y}_{t+h|t}^1$ and $\hat{\varepsilon}_{t+h|t}^2 = y_{t+h} - \hat{y}_{t+h|t}^2$, where $\hat{y}_{t+h|t}^2$ is the forecast from model 2 for h periods ahead for example. The Diebold and Mariano (1995) test is

based on the difference between the forecasting error of the models.

The null hypothesis of the test is that there is equality of forecasting performance between the two models, that is, the models have statistically equal deviations. On the other hand, the alternative hypothesis (one-sided test) defines that the model used as a reference leads to more accurate forecasts than the other. We need to define the forecasts of one of the models as a benchmark for the test, which we will do based on the results from the MSE. The Diebold and Mariano (1995) test is given by:

²These two sectors together have a 2.3% share in the industrial production index.

$$S = \frac{\bar{d}}{\sqrt{\widehat{Avar}(\bar{d})}} \quad (30)$$

where $\bar{d} = \frac{1}{n} \sum_{t=t_0}^T d_t$, $d_t = g(\hat{\varepsilon}_{t+h|t}^1) - g(\hat{\varepsilon}_{t+h|t}^2)$, and $\widehat{Avar}(\bar{d})$ is an estimate of the asymptotic variance (large samples) of \bar{d} for the sample selected. Thus, the S statistic follows a Student t-distribution (Diebold, Mariano, 1995). We consider the quadratic loss function for the test performed in the present work.

4.1.2 Model Confidence Set

Hansen et al (2011) propose the Model Confidence Set (MCS). The interpretation of MCS is similar to a confidence interval. The advantage of this procedure is that it does not require imposing a benchmark. This procedure allows to obtain a set of models that contain the best model with a certain level of confidence. $MCS_{1-\alpha}$ considers the best model with $100(1 - \alpha)\%$ confidence. The set contains more models if we decrease α .

MCS uses a loss function to create the test statistics. We use squared errors as the loss function. The MCS procedure estimates p-values for all models from this statistics, in which the variance is calculated by the bootstrap estimation. The null hypothesis of MCS is the equal predictive ability between the models of the set.

Consider the loss statistics $\bar{d}_{ij} = n^{-1} \sum_{t=1}^n d_{ij,t}$ where $d_{ij,t} = g(\hat{\varepsilon}_{t+h|t}^i) - g(\hat{\varepsilon}_{t+h|t}^j)$, g as a quadratic function

and we have the sample size n for the prediction errors. \bar{d}_{ij} is the relative loss between the i th and j th models and $\bar{d}_i = m^{-1} \sum_{j \in M} d_{ij}$, where \bar{d}_i is the loss of the i th model relative to the average across models in M .

We obtain the set of models until the null hypothesis is no longer rejected for α that we establish. Hansen et al (2011) present two different statistics: $T_{max,M}$ and $T_{R,M}$. We consider the first statistic because it is simple and easy to compute. The second statistic compares all models two by two to get the set, but the computational process is more intense. We construct the statistics

$$t_i = \frac{\bar{d}_i}{\widehat{var}(\bar{d}_i)}$$

Note that $\widehat{var}(\bar{d}_i)$ is the bootstrapped estimate of $var(\bar{d}_i)$. $T_{max,M} = \max_{i \in M} t_i$, where t_i is the statistics about the sample loss between i th and across models in M . The asymptotic distribution of this statistic is nonstandard and Hansen et al (2011) propose to use bootstrap methods.

Then, the algorithm is a sequential procedure and it starts by stating that the set of models is the total of models considered in the work, that is, $M = M^0$ in the first step. The second step is to test the null hypothesis at level α . The third step is if the null hypothesis is accepted, the final set is $M = M^0$; otherwise we eliminate the model with the lowest p-value from M and we repeat the procedure considering $M = M^1$, where M^1 is M^0 without the worst model by the p-value.

5 Results

This section presents the forecasts of 1 to 12 months ahead for the naive models - AR(1), AR(1) with time-varying parameters, UC with SV - besides the lag selection from the LASSO and its two variants (adaLASSO and WLadaLASSO), exponential smoothing ETS and Autometrics with and without the inclusion of IIS and SIS variables to forecast the general industry from aggregated and disaggregated data. We have four types of forecasting models with Autometrics. The first is 'without outlier' which does not consider any type of outlier in the general model. The second is 'IIS' that allows only the dummy variables to control for outlier. The third is the 'SIS' that allows only the step dummy variables. Finally, 'IIS and SIS' allows all possible break and outlier variables. We compare the results in three parts, based on the MSE, the Diebold and Mariano (1995) test and the MCS to obtain the model with the most accurate forecast.

Table 3 presents the MSE of 1 to 12 months ahead for the different models. We multiply the MSE value by 10^4 to improve the visualization of the results presented in the table 3. We also put in bold and red the lowest MSE per forecast horizon to make it easier to see the table 3. In general, we can analyze that the MSE is smaller for the disaggregated models in relation to the aggregates independent of the forecast horizon, with the exception of AR (1) with time-varying parameters and UC with SV. The model with the lowest MSE for the forecast of one to seven months ahead is disaggregated ETS. Increasing the forecast time horizon mainly from one to three periods ahead does not greatly affect the MSE of the disaggregated ETS. The Autometrics algorithm presents lower MSE for the time horizon of eight to twelve months. The Autometrics algorithm with SIS variables and a significance level of 0.1% for disaggregated data presents lower MSE for the time horizon of 10 to 12 months ahead. The Autometrics algorithm with IIS dummy variables and with significance level of 1% using aggregate data has the lower MSE only for seven months forward. This is the only model with aggregate data that has lower MSE and it is only in one step forward. Also the Autometrics algorithm with SIS variables and significance level of 1% has the lower MSE for eight months forward. The disaggregated ETS has an increase in MSE for the steps forward while the Autometrics algorithm has not a growth in MSE for larger horizons in general.

The difference in MSE between the LASSO models and variants is small mainly using disaggregated data, in which the disaggregated WLadaLASSO performs better than the other disaggregated LASSO-type methods for the different forecast horizons. However, when we estimate the aggregate model, the LASSO model presents better prediction performance in relation to its variants. An interesting point would be to try to find out why WLadaLASSO does not perform better than its variants for aggregate data or why there is this difference of performance between aggregated and disaggregated data. Konzen and Ziegelmann (2016) point out that WLadaLASSO would perform better than LASSO and adaLASSO when the sample is small and the greater the number of lags used but we are not varying these two points.

The MSE difference does not allow to state if the disaggregated ETS has higher accuracy of the forecast statistically in relation to other methods. Therefore, we analyze the results of the Diebold and Mariano (1995) test. In order to perform the test, we establish the disaggregated ETS as the benchmark model because it has lower MSE of 1 to 7 months ahead compared to the rest of the models, indicating forecast power gains. We compare the forecast errors of all models with those of the benchmark, having as null hypothesis the equality of predictive power. Under the null hypothesis, the disaggregated ETS predicts as well as the analyzed model. The alternative hypothesis indicates that the prediction of the disaggregated ETS is statistically better.

Table 4 presents the Diebold and Mariano (1995) statistic and the associated p-value (below the test statistic) for each of the models in relation to the benchmark. We present in bold and red the p-value when we reject the null hypothesis to facilitate the visualization and interpretation of the table 4. From the results of Diebold and Mariano (1995) test, we can consider that the disaggregated ETS presents better statistical performance for forecast in relation to the naive models (AR(1), AR(1) with time-varying parameters, UC with SV) considering aggregate or disaggregated data and aggregate ETS. However, the disaggregated ETS model does not present better accuracy in relation to the LASSO models and their variants or the Autometrics algorithm models with or without dummy variables, regardless of being aggregated or disaggregated.

But only the MSE is not enough to say whether the forecasting difference between the models is statistically significant. We perform the MCS procedure to obtain which models we can consider as the best. Table 5 presents p-values from the MCS and those models in the set of “best” models with 90% of probability are indicated with one asterisk (and are in bold and in red). We select only ETS models (with aggregated and disaggregated data) and the Autometrics algorithm with IIS dummy variables with aggregated data with 1% of significance level for one

	AR(1)		TVP AR(1)		UC-SV		ETS		LASSO		adaLASSO		WLadaLASSO	
	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag
1	7.96	7.68	9.87	10.30	9.18	9.38	1.61	1.42	2.72	2.25	2.93	2.21	2.80	2.20
2	7.93	7.77	9.98	10.46	8.94	9.10	1.71	1.48	2.74	2.36	2.95	2.36	2.94	2.34
3	7.98	7.86	10.75	11.11	9.18	9.31	1.77	1.51	2.79	2.41	2.96	2.42	3.01	2.41
4	8.10	8.02	11.29	11.53	9.46	9.66	2.07	1.73	2.82	2.43	3.02	2.44	3.09	2.42
5	8.12	8.05	11.34	11.56	9.49	9.64	2.23	1.89	2.84	2.47	3.06	2.47	3.17	2.44
6	8.10	8.05	11.43	11.64	9.54	9.70	2.38	2.06	2.85	2.48	3.09	2.49	3.22	2.46
7	8.15	8.11	11.41	11.68	9.51	9.72	2.67	2.31	2.87	2.50	3.10	2.50	3.26	2.48
8	8.16	8.13	11.37	11.64	9.45	9.67	2.93	2.54	2.90	2.53	3.13	2.53	3.32	2.51
9	8.19	8.17	11.14	11.41	9.30	9.56	3.20	2.78	2.95	2.58	3.19	2.58	3.39	2.55
10	8.19	8.17	10.75	11.01	9.06	9.28	3.51	3.08	2.99	2.62	3.21	2.62	3.43	2.59
11	8.19	8.17	10.34	10.56	8.88	9.12	3.83	3.36	3.02	2.66	3.24	2.66	3.46	2.63
12	8.19	8.17	9.94	10.13	8.72	8.86	4.15	3.68	3.05	2.69	3.26	2.69	3.48	2.67

Continuation

	Autometrics (1%)								Autometrics (0.1%)							
	Without outlier		IIS		SIS		IIS and SIS		Without outlier		IIS		SIS		IIS and SIS	
	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag
1	3.66	3.18	1.73	2.56	2.69	2.63	3.47	2.62	3.42	3.21	3.21	2.49	2.69	2.48	3.33	2.81
2	3.46	3.07	2.02	2.77	2.74	2.69	3.46	2.72	3.45	3.12	3.19	2.69	2.68	2.66	3.25	2.96
3	3.34	3.04	2.13	2.90	2.87	2.74	3.51	2.80	3.40	3.12	3.27	2.85	2.74	2.78	3.31	3.01
4	3.32	3.03	2.17	2.93	2.86	2.79	3.53	2.82	3.43	3.11	3.31	2.90	2.79	2.84	3.33	3.05
5	3.23	3.01	2.26	2.90	2.96	2.68	3.66	2.77	3.38	3.11	3.41	2.80	2.93	2.72	3.41	2.90
6	3.14	2.97	2.35	2.80	2.96	2.58	3.65	2.68	3.40	3.07	3.41	2.66	2.89	2.59	3.42	2.76
7	3.09	2.96	2.36	2.72	2.94	2.49	3.61	2.61	3.41	3.05	3.36	2.54	2.85	2.49	3.38	2.65
8	3.03	2.95	2.38	2.67	2.87	2.41	3.55	2.55	3.39	3.05	3.30	2.46	2.77	2.41	3.33	2.56
9	3.02	2.96	2.41	2.65	2.85	2.37	3.49	2.54	3.37	3.08	3.29	2.44	2.78	2.38	3.31	2.54
10	3.01	2.98	2.45	2.63	2.75	2.35	3.47	2.52	3.33	3.09	3.32	2.40	2.71	2.33	3.33	2.50
11	2.98	2.99	2.47	2.62	2.67	2.34	3.44	2.51	3.28	3.10	3.30	2.38	2.62	2.31	3.31	2.48
12	2.96	2.99	2.48	2.62	2.64	2.34	3.45	2.52	3.26	3.11	3.30	2.36	2.59	2.29	3.30	2.47

Table 3: MSE results for forecasting from 1 to 12 months ahead for different models

	AR(1)		TVP AR(1)		UC-SV		ETS	LASSO		adaLASSO		WLadaLASSO	
	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Ag	Disag	Ag	Disag	Ag	Disag
1	-10.38	-10.37	-11.84	-12.00	-10.82	-10.82	-5.99	2.40	3.96	1.56	4.00	0.72	4.11
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	1.00	0.94	1.00	0.76	1.00
2	-9.51	-9.48	-11.05	-11.18	-10.00	-10.00	-4.23	1.66	2.68	1.08	2.72	0.49	2.78
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	1.00	0.86	1.00	0.69	1.00
3	-9.12	-9.06	-10.17	-10.30	-9.29	-9.29	-3.53	1.40	2.25	0.91	2.29	0.41	2.34
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.99	0.82	0.99	0.66	0.99
4	-8.27	-8.21	-9.67	-9.77	-8.68	-8.68	-3.06	1.27	2.05	0.82	2.09	0.37	2.13
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	0.98	0.79	0.98	0.65	0.98
5	-8.03	-7.99	-9.58	-9.70	-8.52	-8.52	-2.87	1.21	1.94	0.78	1.98	0.35	2.02
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.97	0.78	0.98	0.64	0.98
6	-8.20	-8.13	-9.79	-9.82	-8.67	-8.67	-2.71	1.17	1.88	0.76	1.92	0.34	1.95
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.88	0.97	0.78	0.97	0.63	0.97
7	-8.53	-8.42	-10.50	-10.46	-9.19	-9.19	-2.58	1.15	1.83	0.74	1.87	0.33	1.90
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.87	0.97	0.77	0.97	0.63	0.97
8	-8.34	-8.25	-10.50	-10.43	-9.01	-9.01	-2.49	1.11	1.76	0.71	1.81	0.32	1.84
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.87	0.96	0.76	0.96	0.63	0.97
9	-7.87	-7.79	-10.57	-10.47	-8.81	-8.81	-2.38	1.05	1.69	0.68	1.73	0.31	1.76
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.85	0.95	0.75	0.96	0.62	0.96
10	-7.74	-7.65	-10.48	-10.43	-8.69	-8.69	-2.29	1.00	1.59	0.64	1.63	0.29	1.66
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.84	0.94	0.74	0.95	0.61	0.95
11	-7.57	-7.47	-10.45	-10.33	-8.44	-8.44	-2.21	0.93	1.49	0.60	1.52	0.27	1.55
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.82	0.93	0.73	0.94	0.61	0.94
12	-5.45	-5.41	-6.96	-7.02	-6.00	-6.00	-2.09	0.83	1.34	0.54	1.37	0.24	1.40
	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.80	0.91	0.70	0.91	0.60	0.92

Continuation

	Autometrics (1%)								Autometrics (0.1%)							
	Without outlier		IIS		SIS		IIS and SIS		Without outlier		IIS		SIS		IIS and SIS	
	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag	Ag	Disag
1	2.82	2.75	4.80	4.24	4.11	5.60	0.81	4.70	1.55	2.23	1.31	5.46	4.37	5.86	1.33	5.00
	1.00	1.00	1.00	1.00	1.00	1.00	0.79	1.00	0.94	0.99	0.91	1.00	1.00	1.00	0.91	1.00
2	1.98	1.90	3.25	2.82	2.81	3.72	0.56	3.11	1.10	1.55	0.92	3.64	2.99	3.91	0.92	3.37
	0.98	0.97	1.00	1.00	1.00	1.00	0.71	1.00	0.86	0.94	0.82	1.00	1.00	1.00	0.82	1.00
3	1.69	1.61	2.73	2.37	2.35	3.12	0.48	2.61	0.93	1.32	0.78	3.06	2.50	3.28	0.79	2.83
	0.95	0.95	1.00	0.99	0.99	1.00	0.69	1.00	0.82	0.91	0.78	1.00	0.99	1.00	0.79	1.00
4	1.54	1.47	2.42	2.14	2.14	2.83	0.43	2.36	0.85	1.20	0.70	2.78	2.28	2.99	0.71	2.58
	0.94	0.93	0.99	0.98	0.98	1.00	0.67	0.99	0.80	0.89	0.76	1.00	0.99	1.00	0.76	0.99
5	1.46	1.39	2.23	2.00	2.00	2.65	0.40	2.20	0.80	1.13	0.64	2.61	2.13	2.81	0.65	2.42
	0.93	0.92	0.99	0.98	0.98	1.00	0.65	0.99	0.79	0.87	0.74	1.00	0.98	1.00	0.74	0.99
6	1.41	1.34	2.13	1.92	1.91	2.54	0.38	2.11	0.77	1.09	0.62	2.51	2.05	2.70	0.63	2.32
	0.92	0.91	0.98	0.97	0.97	0.99	0.65	0.98	0.78	0.86	0.73	0.99	0.98	1.00	0.73	0.99
7	1.37	1.30	2.05	1.85	1.83	2.46	0.37	2.03	0.75	1.05	0.60	2.42	1.97	2.61	0.60	2.24
	0.91	0.90	0.98	0.97	0.97	0.99	0.64	0.98	0.77	0.85	0.73	0.99	0.98	1.00	0.73	0.99
8	1.33	1.25	1.94	1.77	1.75	2.36	0.35	1.95	0.72	1.02	0.57	2.33	1.89	2.51	0.57	2.15
	0.91	0.89	0.97	0.96	0.96	0.99	0.64	0.97	0.76	0.84	0.72	0.99	0.97	0.99	0.72	0.98
9	1.27	1.20	1.84	1.69	1.68	2.26	0.33	1.86	0.69	0.97	0.54	2.24	1.81	2.41	0.55	2.07
	0.90	0.88	0.97	0.95	0.95	0.99	0.63	0.97	0.76	0.83	0.71	0.99	0.96	0.99	0.71	0.98
10	1.21	1.14	1.75	1.60	1.59	2.14	0.32	1.76	0.66	0.92	0.52	2.12	1.72	2.29	0.52	1.96
	0.89	0.87	0.96	0.94	0.94	0.98	0.62	0.96	0.74	0.82	0.70	0.98	0.96	0.99	0.70	0.97
11	1.13	1.06	1.64	1.50	1.49	2.01	0.30	1.65	0.62	0.86	0.48	1.98	1.61	2.14	0.49	1.84
	0.87	0.86	0.95	0.93	0.93	0.98	0.62	0.95	0.73	0.81	0.69	0.98	0.95	0.98	0.69	0.97
12	1.01	0.95	1.51	1.37	1.36	1.83	0.27	1.51	0.55	0.77	0.44	1.81	1.47	1.95	0.45	1.67
	0.84	0.83	0.93	0.91	0.91	0.97	0.61	0.93	0.71	0.78	0.67	0.96	0.93	0.97	0.67	0.95

Table 4: Diebold and Mariano (1995) test results from 1 to 12 months ahead for different models

and four steps forward. Only the disaggregated ETS remains in the MCS for forecast horizons of 2 and 3 periods forward. The best model to forecast 5 to 11 months ahead is composed mostly of models with disaggregated data. The best model has the ETS models, Autometrics, LASSO, adaLASSO and WLadaLASSO.³ The best model to forecast 12 months ahead includes only the models selected by the Autometrics algorithm. The two best models out of the four in the MCS for 12 periods forward are the aggregates with Autometrics.

Conclusions

The present work seeks to analyze two points about forecasting Brazilian industrial production. The first is to compare different univariate models for selection of lags like LASSO and two variants, exponential smoothing models and Autometrics algorithm. Among these models, which type of model has the best forecast for the Brazilian industrial production. The second point is to consider whether the disaggregated data contribute to predict the aggregate level of industrial production. Basically our result points to a better performance of models that use disaggregated data in relation to aggregated data. The exponential smoothing model with disaggregated data performs better in the forecast from 1 to 7 months ahead. The Autometrics algorithm with impulse and structural break variables offers better forecasting performance from 8 to 12 months ahead. The model selected by the Autometrics algorithm generates forecasts with the lowest MSE for 10 to 12 steps forward specifying a lower significance level (0.1%) and including structural break variables with disaggregated data. The difference in performance between the LASSO and its variants (adaLASSO, WLadaLASSO) is small when we consider the disaggregated data.

Epprech et al (2019) obtain a different result that adaLASSO and Autometrics have similar forecasting performance with small values of relevant variables and when the candidate variables are lower than the number of observations. The Autometrics algorithm only performs better when they have a large value of relevant variables because of the bias from the penalization term in adaLASSO, according to the authors. In our case, the Autometrics algorithm always has a better performance than the LASSO type penalties.

This is an ongoing research. The next steps are to contemplate two additional models for forecast, which is the lag selection with dynamic model averaging/selection framework of Koop and Korobilis (2012) and Raftery et al (2010) and bagged ETS of Bergmeir, Hyndman and Benítez (2016).

References

- Barhoumi, K., Darné, O., and Ferrara, L. (2010), "Are disaggregate data useful for factor analysis in forecasting French GDP", *Journal of Forecasting*, 29, 132-144.
- Barnett, A., Mumtaz, H., and Theodoridis, K. (2014), "Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters", *International Journal of Forecasting*, 30, 129-143.
- Carlos, T., and Marçal, E. (2016), "Forecasting Brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon", *Applied Economics*, 48, 4846-4860.
- Carter, C., and Kohn, P. (2004), "On Gibbs sampling for state space models", *Biometrika*, 81, 541-53.
- Chib, S. (1995), "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, 90, 1313-21.
- Chib, S., and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output", *Journal of the American Statistical Association*, 96, 270-81.

³Only for two forecast horizons we have the UC-SV model in the best model.

	1	2	3	4	5	6	7	8	9	10	11	12
AR1 Ag	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR1 Disag	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TVPAR Ag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TVPAR Disag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UC-SV Ag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UC-SV Disag	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.315*	0.134*	0.066
ETS Ag	0.178*	0.074	0.086	0.125*	0.243*	0.615*	0.599*	0.458*	0.348*	0.205*	0.134*	0.066
ETS Disag	1.000*	1.000*	1.000*	1.000*	1.000*	1.000*	1.000*	0.988*	0.803*	0.264*	0.134*	0.045
LASSO Ag	0.053	0.030	0.019	0.054	0.162*	0.154*	0.092	0.053	0.021	0.007	0.005	0.001
LASSO Disag	0.053	0.031	0.019	0.070	0.243*	0.615*	0.856*	0.845*	0.629*	0.000	0.000	0.000
adaLASSO Ag	0.044	0.021	0.016	0.040	0.117*	0.060	0.038	0.028	0.013	0.004	0.001	0.000
adaLASSO Disag	0.053	0.031	0.019	0.070	0.243*	0.615*	0.672*	0.453*	0.172*	0.184*	0.134*	0.042
WLadaLASSO Ag	0.053	0.021	0.014	0.040	0.102*	0.066	0.045	0.033	0.008	0.002	0.000	0.000
WLadaLASSO Disag	0.053	0.031	0.019	0.070	0.243*	0.615*	0.984*	0.965*	0.697*	0.000	0.000	0.000
Autometrics without outlier Ag (1%)	0.025	0.003	0.000	0.000	0.002	0.018	0.024	0.006	0.001	0.000	0.000	0.000
Autometrics without outlier Disag (1%)	0.007	0.003	0.001	0.003	0.013	0.010	0.006	0.005	0.001	0.000	0.000	0.000
Autometrics IIS Ag (1%)	0.187*	0.067	0.060	0.157*	0.305*	0.615*	0.984*	1.000*	0.988*	0.879*	0.743*	0.468*
Autometrics IIS Disag (1%)	0.053	0.031	0.019	0.040	0.162*	0.570*	0.588*	0.267*	0.108*	0.083	0.043	0.031
Autometrics SIS Ag (1%)	0.030	0.004	0.006	0.029	0.132*	0.263*	0.225*	0.091	0.039	0.083	0.125*	0.066
Autometrics SIS Disag (1%)	0.037	0.010	0.006	0.021	0.162*	0.615*	0.984*	0.993*	1.000*	0.184*	0.134*	0.066
Autometrics SIS and IIS Ag (1%)	0.030	0.004	0.002	0.001	0.001	0.000	0.000	0.001	0.000	0.001	0.001	0.001
Autometrics SIS and IIS Disag (1%)	0.053	0.022	0.016	0.040	0.162*	0.615*	0.729*	0.474*	0.267*	0.473*	0.278*	0.144*
Autometrics without outlier Ag (0.1%)	0.038	0.003	0.000	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Autometrics without outlier Disag (0.1%)	0.009	0.004	0.002	0.003	0.003	0.002	0.001	0.001	0.000	0.000	0.000	0.000
Autometrics IIS Ag (0.1%)	0.053	0.010	0.006	0.008	0.010	0.005	0.005	0.008	0.008	0.005	0.009	0.008
Autometrics IIS Disag (0.1%)	0.053	0.011	0.006	0.008	0.102*	0.572*	0.729*	0.707*	0.583*	0.315*	0.199*	0.130*
Autometrics SIS Ag (0.1%)	0.053	0.010	0.006	0.008	0.103*	0.615*	0.984*	0.993*	0.988*	1.000*	1.000*	1.000*
Autometrics SIS Disag (0.1%)	0.030	0.004	0.003	0.005	0.062	0.107*	0.066	0.029	0.008	0.004	0.003	0.002
Autometrics SIS and IIS Ag (0.1%)	0.038	0.004	0.006	0.029	0.132*	0.408*	0.318*	0.170*	0.054	0.088	0.134*	0.066
Autometrics SIS and IIS Disag (0.1%)	0.040	0.009	0.006	0.008	0.010	0.004	0.005	0.006	0.006	0.004	0.009	0.009

Table 5: Model Confidence set p-values from 1 to 12 months ahead for different models

- Crone, S., and Kourentzes, N. (2010), "Feature selection for time series prediction – a combined filter and wrapper approach for neural networks", *Neurocomputing*, 73, 1923-1936.
- Duarte, C., and Rua, A. (2007), "Forecasting through a Bottom-up approach: how bottom is bottom?", *Economic Modelling*, 24, 941-953.
- Epprecht, C., Guegan, D., Veiga, A., Rosa, J. (2019), "Variable selection and forecasting via automated methods for linear models: LASSO/adaLASSO and Autometrics", *Communications in Statistics - Simulation and Computation*, forthcoming.
- Espasa, A., Senra, E., and Albacete, R. (2002), "Forecasting inflation in the European Monetary Union: a disaggregated approach by countries and by sectors", *European Journal of Finance*, 8, 402-421.
- Giacomini, R., and Granger, C. W. J. (2004), "Aggregation of space-time processes", *Journal of Econometrics*, 118, 7-26.
- Granger, C. W. J. (1987), "Implications of aggregation with common factors", *Econometric Theory*, 3, 208-222.
- Hansen, P., Lunde, A., and Nason, J. (2011), "The Model Confidence Set", *Econometrica*, 79(2), 453-497.
- Harvey, D., S. Leybourne, and P. Newbold. (1997), "Testing de equality of prediction mean squared errors", *International Journal of Forecasting*, 13, 281-291.
- Hendry, D. F., and Hubrich, K. (2011), "Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate", *Journal of Business and Economic Statistics*, 29, 216-227.
- Hyndman, R., Koehler, A., Snyder, R., and Grose, S. (2002), "A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods", *International Journal of Forecasting*, 18, 439-454.
- Hyndman, R., Koehler, A., Ord, J., and Snyder, R. (2005), "Prediction Intervals for Exponential Smoothing Using Two New Classes of State Space Models", *Journal of Forecasting*, 24, 17-37.
- Hyndman, R., and Khandakar, Y. (2008), "Automatic Time Series Forecasting: the forecast package for R", *Journal of Statistical Software*, 27, 1-22.
- Hyndman, R., Koehler, A., Ord, J., and Snyder, R. (2008), *Forecasting with exponential smoothing*, Springer, Berlin.
- Hubrich, K. (2005), "Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?", *International Journal of Forecasting*, 21, 119-136.
- Jacquier, E, Polson, N, and Rossi, P (2004), "Bayesian analysis of stochastic volatility models", *Journal of Business and Economic Statistics*, 12, 371-418.
- Kapetanios, G., Marcellino, M., and Venditti, F. (2017), "Large time-varying parameter VARs: a non-parametric approach", Bank of Italy working paper no. 1122.
- Koch, A, Teriçoesvirta, T. (2014), "Forecasting performances of three automated modelling techniques during the economic crisis 2007-2009", *International Journal of Forecasting*, 30, 616-631.
- Koop, G., and Korobilis, D. (2012), "Forecasting Inflation Using Dynamic Model Averaging", *International Economic Review*, 53(3), 867-886.
- Kourentzes, N., Barrow, D., and Crone, S. (2014), "Neural network ensembles operators for time series forecasting", *Expert Systems with Applications*, 41, 4235-4244.
- Kroese, D., and Chan, J. (2014), *Statistical Modeling and Computation*, Springer, Berlin.
- Liçetkepohl, H. (1984), "Linear transformation of vector ARMA processes", *Journal of Econometrics*, 26, 283-293.
- Liçetkepohl, H. (1987), *Forecasting Aggregated Vector ARMA Processes*, Springer-Verlag, Berlin.
- Marcellino, M., Stock, J. H., and Watson, M. W. (2003), "Macroeconomic forecasting in the Euro area: Country specific versus area-wide information", *European Economic Review*, 47, 1-18.
- Ord, J., Koehler, A., and Snyder, R. (1997), "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models", *Journal of the American Statistical Association*, 92, 1621-1629.

- Raftery, A., Karny, M., and Ettler, P. (2010), "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill", *Technometrics*, 52(1), 52–66.
- Stock, J., and Watson, M. (2007), "Why has U.S. inflation become harder to forecast?", *Journal of Money, Credit and Banking*, 39, 3-33.
- Tibshirani R. (1996), "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society, Series B*, 58, 267–288.
- Weber, E. and Zika, G. (2016), "Labour market forecasting in Germany: is disaggregation useful?", *Applied Economics*, 48, 2183-2198.
- Zhao, P., and Yu, B. (2006), "On model selection consistency of Lasso", *Journal of Machine learning research*, 7, 2541-2563.

