Economics Bulletin

Volume 40, Issue 3

Nowcasting GDP growth using data reduction methods: Evidence for the French economy

Olivier Darne
LEMNA, University of Nantes

Amelie Charles
Audencia Business School

Abstract

In this paper, we propose bridge models to nowcast French gross domestic product (GDP) quarterly growth rate. The bridge models, allowing economic interpretations, are specified by using a machine learning approach via Lasso-based regressions and by an econometric approach based on an automatic general-to-specific procedure. These approaches allow to select explanatory variables among a large data set of soft data. A recursive forecast study is carried out to assess the forecasting performance. It turns out that the bridge models constructed using the both variable-selection approaches outperform benchmark models and give similar performance in the out-of-sample forecasting exercise. Finally, the combined forecasts of these both approaches display interesting forecasting performance.

Citation: Olivier Darne and Amelie Charles, (2020) "Nowcasting GDP growth using data reduction methods: Evidence for the French economy", *Economics Bulletin*, Volume 40, Issue 3, pages 2431-2439
Contact: Olivier Darne - olivier.darne@univ-nantes.fr, Amelie Charles - acharles@audencia.com.
Submitted: July 07, 2020.   **Published:** September 24, 2020.

1. Introduction

Policy-makers and analysts are continually assessing the state of the economy. However, gross domestic product (GDP) is only available on a quarterly basis with a time span of around 30 days after the end of the reference quarter for France. In this respect, governments and central banks need to have an accurate and timely assessment of GDP growth rate for the current quarter in order to provide a better and earlier analysis of the economic situation.

To provide an economic interpretation to the forecasts, bridge models are an interesting approach for short-term forecasting of real activity. These linear regressions 'bridge' (i.e., link) monthly variables and quarterly GDP growth, allowing to give some explanations on the basis of the evolution of the explanatory indicators. Such models have been widely considered in the literature, especially to forecast GDP growth in national and international institutions (e.g., Baffigi et al., 2004; Golinelli and Parigi, 2007; Barhoumi et al., 2012).[1] In this respect, we propose bridge models to nowcast French quarterly GDP growth, based on a large set of soft data. Survey data or soft data provide a signal that is obtained directly from economic actors and that reects the short-term prospects of their own activity. Further, they are published with very short publications lags and are subject to only very minor corrections.[2]

In recent years, considerable attention has focused on the forecasting of macroeconomic variables in a data-rich environment via the implementation of a variety of machine learning, variable selection and shrinkage methods (e.g., Kim and Swanson, 2014, 2018; Li and Chen, 2014; Veillon, 2019). In this paper, we use linear bridge models explaining French GDP growth for which input variables are selected through a machine learning approach via penalized regression methods and by an econometric approach based on an automatic model selection procedure. We focus on penalized regression methods, which is a generalization of ordinary least squares estimation, with an additional term that penalizes the size of regression coefficients. In doing so, it regularizes the model complexity, and avoids over-fitting that can cause the out-of-sample forecasting performance to deteriorate. For that, we employ Lasso-based regressions, namely Lasso, Adaptive Lasso and Elastic Net methods. The advantages of these penalized regressions are predictive accuracy and model interpretability. For the econometric method we use the automatic model selection procedure based on a general-to-specific (gets) modelling strategy that allows the econometrician to exploit the availability of a large number of data (Barhoumi et al., 2012; Brunhes-Lesage and Darné, 2012). As shown by Castle (2005), gets strategy is appropriate when there is a desire to conform to economic interpretation.

[1] Another way to link a quarterly variable to monthly indicators is the mixed-data sampling (MIDAS) framework. Clements and Galvão (2008) compared MIDAS and bridge equation approaches and found that their performance is comparable.

[2] Alternative data such as Google search data are used by practitioners for short-term macroeconomic forecasting and nowcasting purposes. Nevertheless, the value of these new sources appears to be limited when it comes to analysing the French outlook (Bortoli and Combes, 2015; Bortoli et al., 2018).

The remainder of the paper is organized as follows. Section 2 briefly describes the methodology of the Lasso-based regressions and automated gets procedure. Section 3 presents the results of the variable selections and the out-of-sample forecasting exercise. Section 4 concludes.

2. Methodology

2.1. Penalized regressions

We estimate several shrinkage estimators for linear models with the estimated coefficients given by

$$\widehat{\beta} = \arg \min_{\beta_0,\ldots,\beta_p} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^{p} g(\beta_j;\omega_j;\alpha)$$

where $g(\beta_j;\omega_j;\alpha)$ is a penalty function that depends on the penalty parameter (or tuning parameter) λ and on a weight $\omega_j > 0$. We consider dierent choices for the penalty functions as described below.

Lasso regression: Lasso (*least absolute shrinkage and selection operator*) regression method, which was introduced by Tibshirani (1996), is likely the most well known penalized or regularized regression method. Lasso regression is characterized by an L_1 penalty function, allowing for sparsity:

$$\lambda \sum_{j=1}^{p} g(\beta_j;\omega_j;\alpha) = \lambda \sum_{j=1}^{p} |\beta_j|$$

Lasso regressions tend to produce estimated regression coefficients that are exactly zeros, and thus can be used for variable selections, where only predictors with nonzero estimates are considered to be important (sparse solution). However, Lasso regression has undesirable properties when T is greater than N or when there is a group of variables amongst which all pairwise correlations are very high (Zou and Hastie, 2005). Otherwise, Lasso can generate a selection bias between highly correlated variables.

Adaptive Lasso regression: Zhou (2006) proposes adaptive Lasso (aLasso) to solve the variable selection problem of Lasso, showing that the inclusion of some additional information regarding the importance of each variable could considerably improve the results. aLasso uses the same penalty as Lasso with the inclusion of a weighting parameter that come from a first-step model that can be Lasso, Ridge or even OLS:

$$\lambda \sum_{j=1}^{p} g(\beta_j;\omega_j;\alpha) = \lambda \sum_{j=1}^{p} \omega_j |\beta_j|$$

where $\omega_j = |\beta_j^*|^{-\gamma}$, with $\gamma > 0$ and β_j^* are the coefficients from the rst-step model.

Elastic Net: Zou and Hastie (2005) develop a regularization method, Elastic Net (EN), that is a generalization including Lasso and Ridge as special cases. The EN method simultaneously carries out automatic variable selection and continuous shrinkage, via a (convex) combination of both L_1 and L_2 penalty functions:

$$\lambda \sum_{j=1}^{p} g(\beta_j; \omega_j; \alpha) = \alpha\lambda \sum_{j=1}^{p} \beta_j^2 + (1-\alpha)\lambda \sum_{j=1}^{p} \omega_j |\beta_j|$$

where $\alpha \in [0;1]$. If $\alpha = 0$ or $\alpha = 1$ the EN regression is equivalent to Ridge or Lasso regression, respectively.

The tuning parameter λ in Lasso and aLasso regressions and the two tuning parameter γ and λ in the EN method are selected via cross-validation, which is a data-driven method that is designed to maximize the expected out-of-sample predictive accuracy.

2.2. Automated gets procedure

The automatic model selection procedure is based on a general-to-specific (gets) modelling strategy, proposed by David Hendry.[3]
Autometrics is the computer implementation used in PcGive (Doornik and Hendry, 2018) and based on the automated gets strategy. This automatic model selection procedure has three basic stages in its approach to select a parsimonious undominated representation of an overly general initial model, denoted the general unrestricted model (GUM) containing all variables likely (or specified) to be relevant, including the maximum lag length of the independent and dependent variables: (i) Formulate the GUM that passes a set of chosen diagnostic tests.[4] Each non-significant regressor in the GUM constitutes the starting point of a backward elimination path, and a regressor is non-significant if the p-value of the two-sided t-test is lower than the chosen significance level α (5% by default) (pre-search process); (ii) Undertake backward elimination along multiple path by removing, one-by-one, non-significant regressors as determined by the chosen significance level α. Each removal is checked for validity against the chosen set of diagnostic tests, and for parsimonious encompassing against the GUM; and (iii) Select, among the terminal models, the specification with the best fit according to the BIC.

[3]An overview of the literature, and the developments leading to Gets modelling in particular, is provided by Campos, Ericsson and Hendry (2005).

[4]The statistic tests are the Jarque-Bera normality test, and the Ljung-Box serial correlation test on the residual and the squared residuals.

3. Empirical results

The data set includes large set of monthly soft data, i.e. 52 variables, covering the period of 2003M1 to 2019M12. These data are based on the surveys in manufactured industry, services and construction sector of the Banque de France (BdF) and the French national statistical institute (INSEE) as well as the consumer surveys of the INSEE.

3.1. Variable selections

The results of variable selections are given in Table 1 for the both approaches: penalized regression methods (Lasso, Elastic-Net and aLasso) and econometric approach (Autometrics). The three Lasso-based regressions select the same three variables with slightly different coefficients across the regressions, namely changes in deliveries, changes in orders and changes in production from the survey in manufacturing industry of the Banque de France. As these three variables are highly correlated with coefficients up to 0.95 (Table 2), suggesting a strong collinearity, we keep the variable with the highest coefficient, namely changes in deliveries. Autometrics select only one variable with changes in orders.[5] Note that both approaches select the same variable (changes in orders) and that the automated gets procedure seems to be more parsimonious than the Lasso-based regressions.

We then construct three bridge models (BMs): (i) BM1 based on selection from Autometrics, namely with changes in orders; (ii) BM2 based on selection from Lasso-based regressions, namely with changes in deliveries; and (iii) BM3 based on the third variable selected by the regularized regressions, namely changes in production. These BMs are estimated by ordinary least squares method over the period from 2003Q1 to 2013Q4 using quarterly averages of monthly data as explanatory variables. Various residual diagnostic tests reveal no discernible specification errors.

3.2. Forecasting

Pseudo out-of-sample recursive forecasts are carried out to determine the final equations. The recursive forecasts have been implemented over the period 2014Q1 to 2019Q4 for pseudo out-of-sample. Parameters are estimated at each step but the specification of the models is unchanged. We compare the BMs with two usual univariate benchmarks, namely, naive (NA) and order-one autoregressive (AR1) models, and also a dynamic factor model (DFM) based on the approach of Stock and Watson (2002). We

[5] We compared two model strategies: *Liberal* and *Conservative*, i.e., "target size", which means "the proportion of irrelevant variables that survives the simplification process" (Doornik, 2009) was set to 5% (*Liberal*) and 1% (*Conservative*). The variable selections were similar for both strategies. Epprecht et al. (2019) compared Lasso and aLasso to Autometrics from a simulation experiment and genomic data.

also propose to combine the forecasts of the three BMs (BMcomb) using equal-weight combination (i.e. average forecasts).

For each quarter t, we provide nowcasts of the GDP, $\widehat{Y}_t$, which are obtained from the BMs estimated by OLS. To assess the predictive accuracy, we use the classical mean squared error (MSE) criterion dened by the following equation

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2$$

where $n = 24$ is the number of quarters considered in the recursive forecast exercise, and Y_t is the value – as realized by the French statistical institute – of the macroeconomic variable that we aim to forecast for the quarter t. Obviously, simply comparing MSE-values does not take into account the sample uncertainty underlying observed forecast differences. This is why we additionally applied the testing procedure based on equal predictive ability (EPA) proposed by Harvey et al. (1997) with the modified Diebold-Mariano (MDM) test, which allows for pairwise comparison of forecast performances across models.

Finally, we employ the model condence set (MCS) procedure proposed by Hansen et al. (2011) to determine the set, $\mathcal{M}^*$, that consists of a subset of equivalent models in terms of predictive ability which are superior to the other competing models from a collection of models, $\mathcal{M}_0$. The MCS procedure yields a model condence set, $\widehat{\mathcal{M}}^*$, that is a set of models constructed to contain the best models with a given level of condence. This MCS allows getting several models displaying equivalent forecasting performance and therefore gives robustness to the forecasting exercise rather than to base the forecasting analysis only on one model. The t-statistic is defined as

$$T_{\max\mathcal{M}} = \max_{i\in\mathcal{M}} t_i \qquad \text{with } t_i = \frac{\bar{d}_i}{\sqrt{\widehat{var}(\bar{d}_i)}} \tag{1}$$

where $\widehat{var}(\bar{d}_i)$ denotes the estimate of $var(\bar{d}_i)$, $\bar{d}_i = m^{-1}\sum_{j\in\mathcal{M}}\bar{d}_{ij}$, and $\bar{d}_{ij} = n^{-1}\sum_{t=1}^{n} d_{ij,t}$, with $d_{ij,t} = L_{i,t} - L_{j,t}$ for all $i, j \in \mathcal{M}_0$, and $L_{i,t}$ is a loss function (here MSE).[6] The t-statistic is associated with the null hypothesis $H_{0,\mathcal{M}} : E(\bar{d}_i) = 0$ for all $i \in \mathcal{M}$, where $\mathcal{M} \subset \mathcal{M}_0$. If $H_{0,\mathcal{M}}$ is accepted at level α then the MCS is the set $\widehat{\mathcal{M}}^*_{1-\alpha}$.[7]

Results in terms of MSE are presented in Table 3 as well as the ratio between each MSE with that obtained from naive (ratio1), AR(1) (ratio2) and DFM (ratio3) models with the associated MDM statistic, and the MCS p-values with its rank.

The results in terms of MSE show that (i) the BMs produce MSE lower than those of the two benchmarks and the DFM, (ii) the MSE of the BMs are very close. This result

[6] $\bar{d}_{ij}$ measures the relative sample loss between the i-th and j-th models, while $\bar{d}_i$ is the sample loss of the i-th model relative to the average across the models in $\mathcal{M}$.

[7] The MCS p-values are calculated using bootstrap implementation with 10,000 resamples (Hansen et al., 2011).

is conrmed by the MDM test because all the BMs are significantly better than both benchmarks and, further, they are in $\widehat{\mathcal{M}}^*_{90\%}$. The smallest MSE are given by the BM1 which are also signicantly lower than those of the DFM model. Finally, two models, BM1 and BM2, display MCS p-values up to 0.50, where the BM1 can be considered as the best model since it is the only one with MCS p-values up to 0.90 (rank 1). This finding show that Lasso-based regressions and automatic gets approach give similar performance in the out-of-sample forecasting exercise and can be complementary alternatives in variable selections.

4. Conclusion

This study proposed bridge models to nowcast French GDP quarterly growth rate, using two variable selection approaches on a large data set of soft data, namely Lasso-based regressions and automated general-to-specific procedure. A recursive forecast study was carried out to assess the forecasting performance and showed that the bridge models outperform benchmark models and give similar performance in the out-of-sample forecasting exercise. We also found that the combined forecasts of the Lasso-based regressions and automated gets procedure display interesting forecasting performance. These results suggest that these alternatives in variable selections can be complementary.

In this paper the bridge models are designed to be used on a quarterly basis since we use quarterly averages of monthly data as explanatory variables, without missing values. Further research would be to design the models on a monthly basis to provide three monthly forecasting exercises of GDP growth rate for a given quarter by dealing with the problem of missing values. For the machine learning approach we used the penalized regression methods. We could also consider other approaches for variable selections, such as random forest, in future research.

References

Baffigi, A., Golinelli, R. and G. Parigi (2004) “Bridge model to forecast the euro area GDP” *International Journal of Forecasting* **20**, 447-60.

Barhoumi, K., Darné, O., Ferrara, L. and B. Pluyaud (2012) “Monthly GDP forecasting using bridge models: Application for the French economy” *Bulletin of Economic Research* **64**, 53-70.

Bortoli, C. and S. Combes (2015) “Input from Google Trends for forecasting the short-term economic outlook in France: Limited avenues” *Conjoncture in France, Insee* **March**, 43-56.

Bortoli, C., Combes, S. and T. Renault (2018) “Nowcasting GDP growth by reading newspapers” *Economics and Statistics* **505-506**, 17-33.

Brunhes-Lesage, V. and O. Darné (2012) “Nowcasting the French index of industrial production: A comparison from bridge and factor models” *Economic Modelling* **29**, 2174-2182.
Campos, J., Hendry, D.F. and N.R. Ericsson (2005). *General-to-Specic Modeling*. Vol.1 and 2. Edward Elgar Publishing: Cheltenham.
Castle, J.L. (2005) “Evaluating PcGets and RETINA as automatic model selection algorithms” *Oxford Bulletin of Economics and Statistics* **67**, 837-80.
Clements, M., Galvão A. (2008) “Macroeconomic forecasting with mixed frequency data: Forecasting output growth in the United States” *Journal of Business and Economic Statistics* **26**, 546-554.
Doornik, J.A. (2009) “Autometrics". In J.L. Castle, N. Shephard (eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 88-121. Oxford University Press: Oxford.
Doornik, J.A. and D.F. Hendry (2018) *Empirical Econometric Modelling - PcGive^{TM} 15*, Vol.1, Timberlake.
Epprecht, C., Guégan, D., Veiga, A. 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*, forthcoming.
Golinelli, R. and G. Parigi (2007) “The use of monthly indicators to forecast quarterly GDP in the short run: An application to the G7 countries” *Journal of Forecasting* **26**, 77-94.
Hansen, P.R., Lunde, A. and J.M. Nason (2011) “Model condence sets for forecasting models” *Econometrica* **79**, 453-497.
Harvey, D., Leybourne, S. and P. Newbold (1997) “Testing the equality of prediction mean squared errors” *International Journal of Forecasting* **13**, 281-91.
Kim, H.H. and N.R. Swanson (2014) “Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence” *Journal of Econometrics* **178**, 352-367.
Kim, H.H. and N.R. Swanson (2018) “Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods” *International Journal of Forecasting* **34**, 339-354.
Li, J. and W. Chen (2014) “Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models” *International Journal of Forecasting* **30**, 996-1015.
Stock, J.H. and M.W. Watson (2002) “Macroeconomic forecasting using diffusion indexes” *Journal of Business and Economic Statistics* **20**, 147-162.
Tibshirani, R. (1996) “Regression shrinkage and selection via the Lasso” *Journal of the Royal Statistical Society. Series B* **58**, 267-288.
Veillon, P-A. (2019) “Continuous forecasting of French economic growth: Testing different models of machine learning” *Conjoncture in France, Insee* **December**, 19-32.

Zou, H. (2006) “The adaptive lasso and its oracle properties” *Journal of the American Statistical Association* **101**, 1418-1429.
Zou, H. and T. Hastie (2005) “Regularization and variable selection via the elastic net” *Journal of the Royal Statistical Society, Series B* **67**, 301-320.

Table 1: Results of variable selections.

Series	Lasso	Elastic-Net	aLasso	Autometrics
changes-in-orders	0.0153	0.0170	0.0174	0.0578
changes-in-deliv	0.0372	0.0265	0.0376	
changes-in-prod	0.0009	0.0117	0.0037	

Table 2: Results of correlations.

Series	changes-in-orders	changes-in-deliv	changes-in-prod
changes-in-orders	1.00	0.98	0.95
changes-in-deliv		1.00	0.98
changes-in-prod			1.00

Table 3: Nowcast performance - 2014Q1-2019Q4.

Series	MSE	ratio1	ratio2	ratio3	MCS	rank
BM1	0.052	0.41^{*}	0.56^{*}	0.67^{**}	1.000	1
BM2	0.054	0.42^{*}	0.58^{**}	0.70	0.648	3
BM3	0.057	0.44^{*}	0.61^{**}	0.73	0.395	4
BMcomb	0.052	0.40^{*}	0.56^{*}	0.67^{**}	0.752	2
DFM	0.078	0.60^{**}	0.84	–	0.047	6
NA	0.094	–	1.38	1.65	0.023	7
AR1	0.129	0.72^{*}	–	1.20	0.052	5

Notes: Ratio1, ratio2 and ratio3 are computed as the ratios between each MSE with that obtained from the naive, AR(1) and DFM models, respectively. * and ** significant at 5% and 10% level, respectively, the null of the modified Diebold-Mariano (MDM) test. The alternative of the MDM test is that the first forecast (naive, AR(1) or DFM models) is less accurate than the second forecast.