



HAL
open science

Are national or regional surveys useful for nowcasting regional jobseekers ?

Clément Cariou, Amélie Charles, Olivier Darné

► To cite this version:

Clément Cariou, Amélie Charles, Olivier Darné. Are national or regional surveys useful for nowcasting regional jobseekers ?. *Journal of Forecasting*, In press, 43 (6), pp.2341-2357. 10.1002/for.3125 . hal-04573788v2

HAL Id: hal-04573788

<https://hal.science/hal-04573788v2>

Submitted on 22 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Are national or regional surveys useful for nowcasting regional jobseekers? The case of the French region of Pays-de-la-Loire

Clément Cariou¹ | Amélie Charles² | Olivier Darné¹

¹LEMNA, Nantes University, Nantes, BP, France

²Audencia Business School, Nantes Cedex 3, France

Correspondence

LEMNA, Nantes University, Chemin de la Censive du Tertre, BP 52231, 44322 Nantes, France.

Email: clement.cariou@univ-nantes.fr

Abstract

In this paper we develop nowcasting models for the Pays-de-la-Loire's jobseekers, a dynamic French regional economy. We ask whether these regional nowcasts are more accurate by only using the regional data or by combining the national and regional data. For this purpose, we use penalized regressions, random forest, and dynamic factor models as well as dimension reduction approaches. The best nowcasting performance is provided by the DFM estimated on the regional and regional-national databases as well as the Elastic-Net model with a prior screening step for which the national data are the most frequently selected data. For the latter, it appears that the Change in foreign orders in the industry sector, the OECD Composite leading indicator, and the BdF Business sentiment indicator are among the major predictors.

KEYWORDS

factor model, jobseekers, nowcasting, penalized regression, regional data, variable selection

1 | INTRODUCTION

Unemployment is a crucial issue for policy-makers. There are societal, governmental, and economic costs of high unemployment. Unemployment diminishes the disposable income of families, household consumption, purchasing power, the employee morale, among others, and increases public spending in terms of unemployment benefits and active labor market policies. It is also considered a major macroeconomic indicator for analysts to have an accurate assessment of the state of the labor market. In France, the regions are in charge of the coordination on their territory of all the actions in favor of the economy. For instance, they are in charge of the professional training of young people and job seekers.

Therefore, early and accurate forecasts are required to propose the appropriate economic and welfare policies at both national and regional levels.

The French National Statistical Institute (INSEE) publishes the unemployment rate on a quarterly basis, with a significant delay of around 45 days after the end of the reference quarter, whereas *Pôle Emploi*, the National Employment Agency in charge of allocating unemployment benefits and job placement, publishes the number of jobseekers at the end of the month (DEFM, *demandeurs d'emploi en fin de mois*) on a monthly basis, with a lag of around 26 days after the end of the reference month. In this paper, we focus on nowcasting the jobseekers of the Pays-de-la-Loire (PID hereafter), a region of western France, which is a regional dynamic economy

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Journal of Forecasting* published by John Wiley & Sons Ltd.

displaying the second lowest level of unemployment for the 2022Q2 with 5.9%, largely under the national level with 7.4%.¹ The PdL's employment dynamism can be explained by the structure of economic activity, the population growth, the economic performance of companies, the number of business creations (companies and micro-enterprises), and the attractiveness of the region. Figure 1 shows that French's and PdL's jobseeker growths seem to display similar behavior.

Traditional unemployment forecasting models are based on survey data because they provide short-term information from economic actors, which are published with a very short delay (e.g., Abberger, 2007; Siliverstovs, 2013; Soybilgen & Yazga, 2018). Further, survey data can be also available at a regional level. Recently, Google search data have been shown to have predictive power for unemployment (e.g., Caperna et al., 2022; Choi & Varian, 2009; D'Amuri & Marcucci, 2017; Fondeur & Karamé, 2013; Nagao et al., 2019; Niesert et al., 2020). Their main advantages are that they are free from revisions and can be obtained in specific geographical areas, such as the region of a country. However, the majority of studies on unemployment forecasting are conducted at the national level and there are very few at the regional level (see, for instance, Simionescu & Cifuentes-Faura, 2022).

The aim of our paper is threefold. First, we nowcast the jobseekers of a specific French region, namely the Pays-de-la-Loire, in a data-rich environment formed by survey data and Google search activity. Second, we question whether it is more appropriate to only use regional data or combine national and regional data by employing various methods applied in the literature for

unemployment forecasting with a large dataset of predictors, i.e. penalized regressions, random forest and dynamic factor model (see, for instance, Kim & Swanson, 2014, 2018; Li & Chen, 2014; Kotchoni et al., 2019; Goulet Coulombe et al., 2022; Gogas et al., 2022). Third, we assess the interest of a targeting step for jobseeker nowcasting by comparing the predictive performance of the various methods with or without an initial variable selection step. Finally, we question whether the predictive performance of the models is different during crisis (Global Financial Crisis and COVID-19 crisis) and non-crisis periods.

The structure of the paper is organized as follows. Section 2 describes the data we consider to nowcast PdL's jobseekers. Section 3 introduces the methodology of the penalized regressions, random forest, dynamic factor model, and screening methods. The nowcasting evaluation is presented in Section 4. In Section 5 we discuss empirical findings, and Section 6 concludes.

2 | DATA

The data set includes monthly data for the number of end-of-month jobseekers from claimant count ("*Demandeurs d'Emploi en Fin de Mois*", category A, DEFM, hereafter) for the region *Pays de la Loire* (PdL, hereafter) provided by *Pôle Emploi*, the National Employment Agency.² This variable describes the inventory of unemployed people at the end of each month. The concept of jobseekers enrolled at *Pôle Emploi* is a different concept to that of unemployment in the sense of the International Labour Office (ILO), and thus some unemployed are not

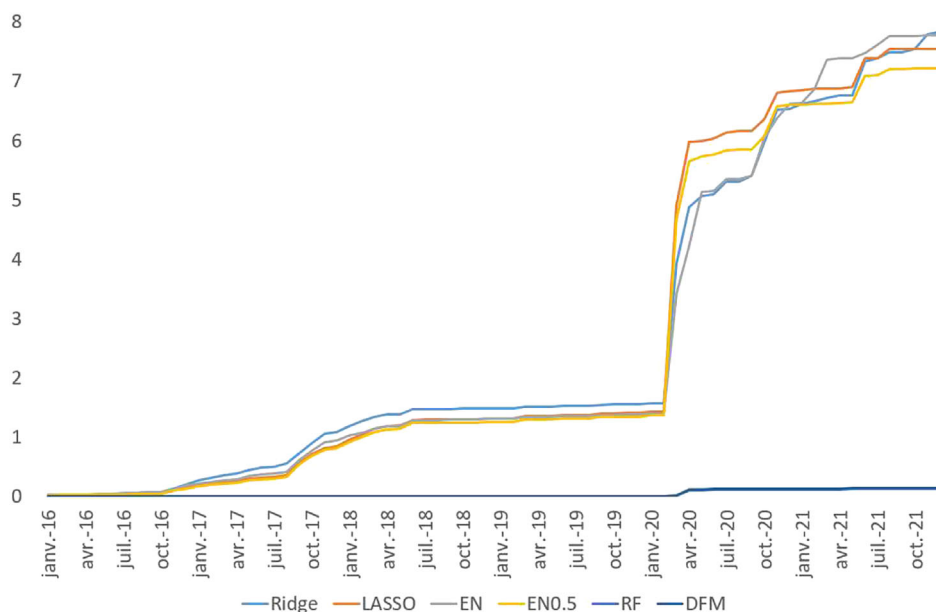
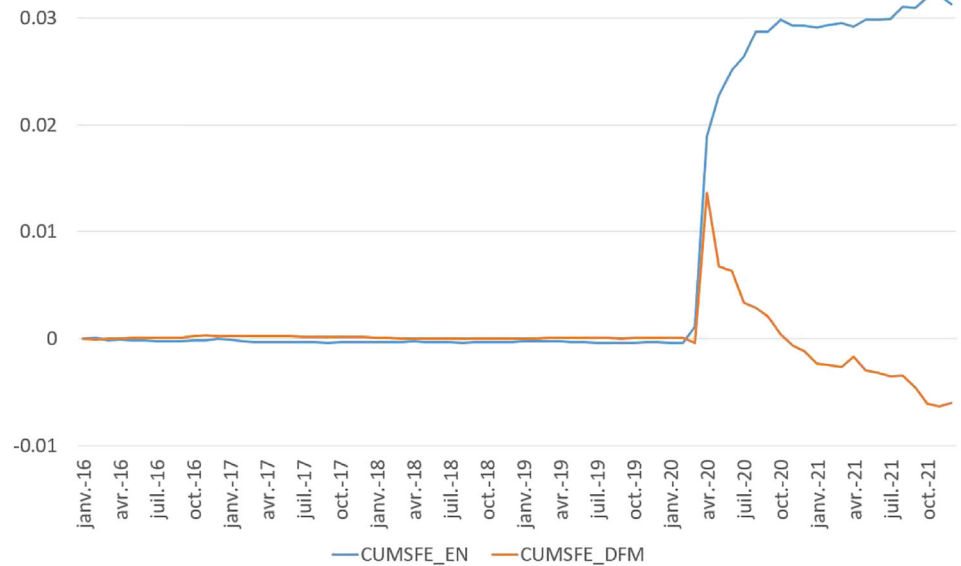


FIGURE 1 Pays-de-la-Loire and French jobseeker growth rate - 2004 M6-2021M12.

FIGURE 2 PdL's jobseeker growth rate - 2004 M6-2021M12.



unemployed in the sense of the ILO, and conversely some unemployed are not enrolled at Pôle Emploi. Data on unemployed people are based on administrative records in center jobs whereas they are based on an employment survey for the ILO. According to the ILO definition, an unemployed person is a person of working age (i.e. aged 15 years or older) who does not work, not even one hour during the week, who is available to take a job within 15 days and who actively sought a job in the previous month. In each country, a statistical survey is conducted to check whether these criteria are met. In France, the INSEE is the national statistical organization responsible for this survey, and it publishes it at a quarterly frequency. End-of-month job seekers are persons registered with *Pôle Emploi* and have an application in progress on the last day of the month. *Pôle Emploi* delivers monthly statistics of jobseekers, following five statistical categories according to the availability of the job applicant. In our study, we focus on the main category, namely category A, which consists of jobseekers required to actively seek employment and unemployed.³ Figure 2 shows that the PdL jobseekers have seen a large rise during the COVID-19 crisis, with a large increase of 32.1% in April 2020.

We consider a large set of monthly indicators selected on the basis of their reliability and timeliness covering the period of June 2004 to December 2021, consisting of business surveys, financial data, and uncertainty indicators. The financial data concern the stock market and short- and long-term interest rates, and we also include the economic policy uncertainty indicator for France proposed by Baker and Fradkin (2017). The business survey data are on the manufacturing and services sectors and are mainly produced by the Banque de France (BdF), the French National Institute of Statistics (INSEE) the IHS

Markit, and also the Organization for Economic Cooperation and Development (OECD) and the European Commission (EC) with business tendency survey indicators.

Google Trends provides information on the search intensity for a particular keyword or topic relative to the total number of searches conducted over time in a given area. Search intensity data are scaled to the highest value over the period of 100. As in Fondeur and Karamé (2013), we use the keyword “jobs” (“*emploi*” in French). We construct two Google Trends variables by specifying, first, France as the geographic zone, and, second, the region Pays-de-la-Loire as the geographic zone.

We construct two different monthly databases, starting from the same data set, which includes the financial data, the economic policy uncertainty indicator, and the Google Trend variables. In the first database, called the regional database, we add the regional surveys provided by the BdF on the manufacturing and service sectors, giving a set of 30 variables. For the second database, called the large database, the national surveys provided by the BdF, the INSEE, the OECD, the EC, and the IHS Markit are added to the regional database, leading to a large database of 134 predictors. We also consider four autoregressive terms. We transform all variables to be stationary. The description of the variables is given in Tables 1–4.

3 | METHODOLOGY

3.1 | Penalized regressions

Tibshirani (1996) suggests the Lasso (*least absolute shrinkage and selection operator*) regression, which is one

TABLE 1 Description of the variables.

Variable	Sector	Code	Source
<i>National level</i>			
Business sentiment indicator	Services	BSI serv BdF	Banque de France
Business sentiment indicator	Industry	BSI indus BdF	Banque de France
Change in employment	Industry	EMP indus BdF	Banque de France
Expected employment	Industry	EEMP indus BdF	Banque de France
Change in employment	Services	EMP serv BdF	Banque de France
Expected employment	Services	EEMP serv BdF	Banque de France
Change in foreign orders	Industry	FORDER indus BdF	Banque de France
Change in overall level of new orders	Industry	ORDER indus BdF	Banque de France
Change in deliveries	Industry	DELIV indus BdF	Banque de France
Change in commodity prices	Industry	CPRICE indus BdF	Banque de France
Current position in inventories of commodities	Industry	INVENT indus BdF	Banque de France
Change in industrial producer prices	Industry	IPRICE indus BdF	Banque de France
Expected industrial prices	Industry	FIPRICE indus BdF	Banque de France
Change in output	Industry	OUTPUT indus BdF	Banque de France
Expected production	Industry	EOUTPUT indus BdF	Banque de France
Change in inventories of final goods	Industry	CINVENTFG indus BdF	Banque de France
Current position in inventories of final goods	Industry	INVENTFG indus BdF	Banque de France
Expected inventories of final goods	Industry	EINVENTFG indus BdF	Banque de France
Cash positions	Industry	CASH indus BdF	Banque de France
Average capacity utilisation rate	Industry	CAPA indus BdF	Banque de France
Current order books	Industry	CORDER indus BDF	Banque de France
Expected overall activity	Services	EACT serv BdF	Banque de France
Change in foreign orders	Services	FORDER serv BdF	Banque de France
Forecasts on foreign orders	Services	FFORDER serv BdF	Banque de France
Change in aggregate demand	Services	DEMAND serv BdF	Banque de France
Expected aggregate demand	Services	EDEMAND serv BdF	Banque de France
Change in prices	Services	PRICE serv BdF	Banque de France
Expected prices	Services	EPRICE serv BdF	Banque de France
Cash positions	Services	CASH serv BdF	Banque de France

Notes: Each variable is at the national level. The variables have been seasonally adjusted with X13-SEATS-ARIMA when necessary.

of the most well-known penalized regression. To achieve sparsity, the Lasso estimator uses an L_1 penalty function as follows

$$\hat{\beta}^{\text{Lasso}} = \underset{\beta_0, \dots, \beta_p}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (1)$$

where $\lambda > 0$ is the tuning parameter. A drawback of the Lasso regression is that it tends to generate a selection bias between highly correlated variables, leading to variable selection problems (arbitrary selection and under-

representation of important variables). We consider two solutions proposed in the literature. The first is the adaptive Lasso (aLasso) regression of Zou (2006):

$$\hat{\beta}^{\text{aLasso}} = \underset{\beta_0, \dots, \beta_p}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j| \right\} \quad (2)$$

where $\hat{\omega}_j = \left| \hat{\beta}_j^* \right|^{-\gamma}$ are weights obtained from a first-step Ridge regression, with $\gamma > 0$. The second approach is the Elastic-Net (EN) regression of Zou and Hastie (2005):

TABLE 2 Description of the variables (continued).

Variable	Sector	Code	Source
<i>National level</i>			
Business climate summary indicator		BCI INSEE	INSEE
Employment climate summary indicator	Employment	ECI INSEE	INSEE
Business climate summary indicator	Industry	BCI indus INSEE	INSEE
Business climate summary indicator	Services	BCI serv INSEE	INSEE
Business climate summary indicator	Construction	BCI const INSEE	INSEE
Business climate summary indicator	Retail	BCI retail INSEE	INSEE
Expected trend in the workforce	Industry	EWORk indus INSEE	INSEE
Past trend in the workshop	Industry	PWORK indus INSEE	INSEE
Level of stocks of manufactured products	Industry	STOCK indus INSEE	INSEE
Past trend in production	Industry	PROD indus INSEE	INSEE
Expected trend in production	Industry	EPROD indus INSEE	INSEE
Probable trend in industrial production volume	Industry	PVOL indus INSEE	INSEE
Probable trend in sale prices	Industry	SPRICE indus INSEE	INSEE
Probable trend in industrial price general level	Industry	IPRICE indus INSEE	INSEE
Past trend of activity	Services	PACTIV serv INSEE	INSEE
Expected trend of activity	Services	EACTIV serv INSEE	INSEE
Expected trend of demand	Services	EDEMAND serv INSEE	INSEE
Past trend of prices	Services	PPRICE serv INSEE	INSEE
Expected trend of prices	Services	EPRICE serv INSEE	INSEE
General prospects on activity	Services	PRACTIV serv INSEE	INSEE
Change in the situation of the enterprise over the last 3 months	Services	ENTERP serv INSEE	INSEE
Expected trend of employment	Services	EEMP serv INSEE	INSEE
Past trend of employment	Services	PEMP serv INSEE	INSEE
Expected trend of employment	Construction	EEMP const INSEE	INSEE
Past trend of employment	Construction	PEMP const INSEE	INSEE
Change in the number of persons employed (next 3 M)	Retail	EMP + 3 retail INSEE	INSEE
Change in the number of persons employed (last 3 M)	Retail	EMP-3 retail INSEE	INSEE
Level of global order books	Industry	ORDER indus INSEE	INSEE
Level of foreign order books	Industry	FORDER indus INSEE	INSEE
Business development (sales) (next 3 M)	Retail	BUSIN retail INSEE	INSEE
Intents for orders (next 3 M)	Retail	ORDER retail INSEE	INSEE
Opinion on the future trend of unemployment	Consumer	UNEMP consum INSEE	INSEE

Notes: Each variable is at the national level. The variables have been seasonally adjusted with X13-SEATS-ARIMA when necessary.

$$\hat{\beta}^{EN} = \underset{\beta_0, \dots, \beta_p}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{i,j} \right)^2 + \lambda \sum_{j=1}^p \left(\alpha \beta_j^2 + (1 - \alpha) |\beta_j| \right) \right\} \quad (3)$$

$\alpha \in [0; 1]$. The EN regression combines the Lasso and Ridge regressions. The tuning parameters λ and α are selected via cross-validation (with 10 folds), and the parameter γ is set to 0.5.

TABLE 3 Description of the variables (continued).

Variable	Sector	Code	Source
<i>National level</i>			
Composite leading indicator		CLI	OECD
Business confidence index		BCI	OECD
Consumer confidence index		CCI	OECD
Economic sentiment indicator		ESI	European Commission
Economic sentiment indicator	Construction	ESI const	European Commission
Economic sentiment indicator	Industry	ESI indus	European Commission
Economic sentiment indicator	Retail	ESI retail	European Commission
Economic sentiment indicator	Consumers	ESI consum	European Commission
Economic sentiment indicator	Services	ESI serv	European Commission
Output	Composite	OUTPUT comp PMI	IHS Markit
New orders	Composite	ORDER comp PMI	IHS Markit
Employment	Composite	EMP comp PMI	IHS Markit
Backlogs of work	Composite	BW comp PMI	IHS Markit
Input prices	Composite	IPRICE comp PMI	IHS Markit
Output prices	Composite	OPRICE comp PMI	IHS Markit
PMI index	Industry	PMI index	IHS Markit
Output	Industry	OUTPUT indus PMI	IHS Markit
New orders	Industry	ORDER indus PMI	IHS Markit
New export orders	Industry	EXORDER indus PMI	IHS Markit
Employment	Industry	EMP indus PMI	IHS Markit
Backlogs of work	Industry	BW indus PMI	IHS Markit
Quantity of purchases	Industry	QPUR indus PMI	IHS Markit
Stocks of purchases	Industry	SPUR indus PMI	IHS Markit
Stocks of finished goods	Industry	SFG indus PMI	IHS Markit
Suppliers' delivery times	Industry	DEILV indus PMI	IHS Markit
Input prices	Industry	IPRICE indus PMI	IHS Markit
Output prices	Industry	OPRICE indus PMI	IHS Markit
Output	Services	OUTPUT serv PMI	IHS Markit
New orders	Services	ORDER serv PMI	IHS Markit
Employment	Services	EMP serv PMI	IHS Markit
Backlogs of work	Services	BW serv PMI	IHS Markit
Input prices	Services	IPRICE serv PMI	IHS Markit
Output prices	Services	OPRICE serv PMI	IHS Markit
Future activity	Services	FACTIV serv PMI	IHS Markit

Notes: The PMI composite index is the weighted average of manufacturing and services sectors. Each variable is at the national level. The variables have been seasonally adjusted with X13-SEATS-ARIMA when necessary.

3.2 | Random forest

Unlike the penalized regressions, random forest (RF), which was introduced by Breiman (2001), allows nonlinearity by using regression trees and bagging.

Let Y_{t+h} be the dependent variable, and X_t a vector of predictors, the tree regression forecast is obtained by

$$Y_{t+h} = \sum_{k=1}^K c_k \mathbf{1}_{(X_t; \theta_k) \in R_k},$$

TABLE 4 Description of the variables (continued).

Variable	Sector	Code	Source
<i>National level</i>			
Consumer price index year-on-year change		CPI	INSEE
CAC40 (return)	Finance	CAC40	Banque de France
Long-term interest rate	Finance	LTIR	Banque de France
Short-term interest rate	Finance	STIR	Banque de France
Spread long-short	Finance	SPREAD	Banque de France
Economic policy uncertainty	Uncertainty	EPU	Baker et al. (2016)
Keyword “emploi” as search term		Google	Google trends
Keyword “emploi” as search term		dGoogle	Google trends
Keyword “emploi” as topic		Google topic	Google trends
Keyword “emploi” as topic		dGoogle topic	Google trends
<i>Regional level</i>			
Jobseekers (growth rate)		DEFM	Pôle Emploi
<i>(Demandeurs d'Emploi en fin de Mois)</i>			
Business sentiment indicator	Services	BSI PdL serv BdF	Banque de France
Business sentiment indicator	Industry	BSI PdL indus BdF	Banque de France
Change in employment	Industry	EMP PdL ind BdF	Banque de France
Expected employment	Industry	PEMP PdL ind BdF	Banque de France
Change in employment	Services	EMP PdL serv BdF	Banque de France
Expected employment	Services	PEMP PdL serv BdF	Banque de France
Change in foreign orders	Industry	FORDER PdL ind BdF	Banque de France
Change in overall level of new orders	Industry	ORDER PdL ind BdF	Banque de France
Change in commodity prices	Industry	CPRICE PdL ind BdF	Banque de France
Change in industrial producer prices	Industry	PPRICE PdL ind BdF	Banque de France
Expected industrial prices	Industry	IPRICE PdL ind BdF	Banque de France
Change in output, compared	Industry	OUTPUT PdL ind BdF	Banque de France
Expected production	Industry	EPROD PdL ind BdF	Banque de France
Change in inventories of final goods	Industry	CINVENTFG PdL ind BdF	Banque de France
Current position in inventories of final goods	Industry	INVENTFG PdL ind BdF	Banque de France
Average capacity utilisation rate	Industry	CAPA PdL ind BdF	Banque de France
Change in activity	Services	ACTIV PdL serv BdF	Banque de France
Expected overall activity	Services	EACTIV PdL serv BdF	Banque de France
Change in prices	Services	PRICE PdL serv BdF	Banque de France
Expected prices	Services	EPRICE PdL serv BdF	Banque de France
Cash positions	Services	CASH PdL serv BdF	Banque de France
Keyword “emploi” as search term		Google PdL	Google trends
Keyword “emploi” as search term		dGoogle PdL	Google trends
Keyword “emploi” as topic		Google topic PdL	Google trends
Keyword “emploi” as topic		dGoogle topic PdL	Google trends

Notes: Each variable is at the regional level. The variables have been seasonally adjusted with X13-SEATS-ARIMA when necessary.

where K is the number of terminal nodes, c_k are node means, i.e. $c_k = \sum_{j \in R_k} Y_j / N_k$, with N_k the number of variables in the k th region, θ_k is the set of parameters defining the k th region, and R_1, \dots, R_K represents the region partition of the space of predictors X_t . The aggregation by averaging of the predictions of each tree gives the final forecast:

$$\hat{Y}_{t+h} = \frac{1}{B} \sum_{b=1}^B Y_{t+h}^b$$

where Y_{t+h}^b represents each regression tree specified on a bootstrapped subsample of the original data, and $b = 1, \dots, B$, with B the number of bootstrap samples.

3.3 | Dynamic factor model

The factor model is based on the decomposition of the variables X into the sum of a common component χ and a idiosyncratic component ξ , which are mutually orthogonal. Given X an $(T \times n)$ matrix of observations, the factor model is given by

$$X = \chi + \xi = F\Lambda' + \xi$$

where ξ is the $(T \times r)$ matrix of idiosyncratic components, Λ is the loading matrix of $(n \times r)$ dimension, and F is the $(T \times r)$ matrix of estimated factors by using the static principal component analysis (PCA) suggested by Stock and Watson (2002),⁴ with $\hat{S} = (\hat{S}_1, \dots, \hat{S}_r)$ the r largest eigenvalues for $j = 1, \dots, r$. An autoregressive structure on the factors is introduced to capture dynamics in forecasting.

3.4 | Dimensionality reduction methods

The dimensionality reduction methods we consider are based on the sure independence screening (SIS) procedure proposed by Fan and Lv (2008). The SIS-type procedures define a submodel $\hat{\mathcal{M}}$ including a set of important predictors exhibiting the largest values of a utility measure computed between the dependent variable Y and each predictor $X = \{X_k\}_{k=1}^p$ as follows

$$\hat{\mathcal{M}} = \{k : \hat{\omega}_k \text{ is among the first } d_n \text{ largest of all, for } 1 \leq k \leq p\}$$

where $\hat{\omega}_k$ is a marginal utility measure, d_n is the submodel size such that $d_n < n$, with n the sample size. The SIS

procedure imposes some strong assumptions on the model and the error distribution, such as linear models and Gaussian errors. We consider two solutions, which are model-free screening procedures. The first is the DC-SIS approach of Li et al. (2012), which is based on the distance correlation (DC) (Székely et al., 2007) as a marginal utility measure, defined by

$$\begin{aligned} \hat{\omega}_k^{dcsis} &= \widehat{\text{dcorr}}(\mathbf{X}_k, \mathbf{Y})^2 \\ &= \frac{\rho \arcsin(\rho) + \sqrt{1-\rho^3} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1}{1 + \pi/3 - \sqrt{3}} \end{aligned}$$

where ρ is the Pearson correlation coefficient. The second alternative is the MDC-SIS procedure of Shao and Zhang, which uses the martingale difference correlation (MDC) as a marginal utility measure, given by

$$\hat{\omega}_k^{mdcsis} = (\text{MDC}_n^j(\mathbf{Y}|\mathbf{X}_k))^2 = \frac{\text{MDD}(\mathbf{Y}|\mathbf{X}_k)^2}{\sqrt{\text{var}^2(\mathbf{Y}) \text{dvar}^2(\mathbf{X}_k)}}$$

where $\text{MDD}(\mathbf{Y}|\mathbf{X}_k)$ denotes the martingale difference divergence (MDD) given by

$$\text{MDD}(\mathbf{Y}|\mathbf{X})^2 = \int_{\mathcal{R}^q} |\psi_{\mathbf{Y},\mathbf{X}}(\mathbf{s}) - \psi_{\mathbf{Y}}\psi_{\mathbf{X}}(\mathbf{s})|^2 \omega(\mathbf{s}) d\mathbf{s}$$

with $\psi_{\mathbf{Y},\mathbf{X}}(\mathbf{s})$, $\psi_{\mathbf{Y}}$, and $\psi_{\mathbf{X}}(\mathbf{s})$ the joint and marginal characteristic functions, and $\omega(\mathbf{s}) = \{c_q |s|_q^{1+q}\}^{-1}$.

Following standard practice in the literature (see, for instance, Fan & Lv, 2008; Li et al., 2012), the tuning parameter d_n is chosen to be $n/\log(n)$ for the two SIS-type procedures.

4 | OUT-OF-SAMPLE NOWCASTING DESIGN

The nowcasting period is January 2016 to December 2021, including the COVID-19 crisis. All models are estimated using a rolling-window scheme, with a fixed rolling window of 138 observations, and the parameters are re-estimated at each step. The first estimation uses thus the first 12-year period, from June 2004 to December 2015 ($T = 138$ observations) to produce the first out-of-sample nowcast, and so on.

The forecasting performance is evaluated on the basis of the mean squared error (MSE) and the out-of-sample (OOS) R^2 , given by $R_{OOS}^2 = 1 - (\text{MSE}_{\text{model}} / \text{MSE}_{\text{benchmark}})$, where $\text{MSE}_{\text{model}}$ and $\text{MSE}_{\text{benchmark}}$ are the MSE of the given and benchmark models (AR model or DFM),

TABLE 5 Most important variables during the full OOS period (2016 M1-2021M12).

Variable	Regional database		Large database					
	Lasso	EN	Lasso	EN	DCSIS+Lasso	DCSIS+EN	MDCSIS+Lasso	MDCSIS-EN
BSI PdL serv BdF					X	X	X	
EMP PdL ind BdF	X	X	X					
EEMP PdL ind BdF	X		X					
FORDER PdL ind BdF	X							
ORDER PdL ind BdF	X	X						
CPRICE PdL ind BdF	X	X						
FIPRICE PdL ind BdF	X	X						
EPROD PdL ind BdF		X		X				
CINVENTFG PdL ind BdF		X						
INVENTFG PdL ind BdF	X	X						
ACTIV PdL serv BdF	X	X	X					
EACTIV PdL serv BdF	X	X						
CLI OECD					X	X	X	X
ESI					X	X	X	
ESI retail				X				
BSI ind BdF					X	X	X	X
EMP serv BdF							X	X
FORDER ind BdF			X	X	X	X	X	X
ORDER ind BdF			X		X			
OUTPUT ind BdF					X			
EPROD ind BdF			X	X				
EACTIV serv BdF			X	X	X		X	X
FORDER serv BdF			X					
DEMAND serv BdF					X	X		
EDEMAND serv BdF					X	X		
BCI INSEE					X			
BCI serv INSEE						X	X	
BCI retail INSEE							X	
EPROD ind INSEE				X				
PACTIV serv INSEE				X				
EDEMAND serv INSEE						X		
EEMP serv INSEE				X				
dGoogle			X	X				
ORDER comp PMI							X	X
BW comp PMI							X	X
PMI index						X		X
ORDER ind PMI								X
BW ind PMI						X		
SFG ind PMI			X	X				
OUTPUT serv PMI						X		X

Notes: The table reports the ten most important variables for each method.

respectively. A positive R_{OOS}^2 means that the given model is more accurate than the benchmark. The modified Diebold-Mariano (MDM) test for a small sample proposed by Harvey et al. (1997) is used to test the predictive accuracy of each model against the benchmark. Its null of equal predictive ability (EPA) means that statistically both models have the same predicting ability, and the alternative is that the given model outperforms the reference.⁶ We also implement the model confidence set (MCS) procedure of Hansen et al. (2011). For a given confidence level α , this procedure constructs a set of superior models (SSM), \mathcal{M}^* , where the null hypothesis of EPA is

not rejected, from a set of competing models, based on a sequential elimination procedure.⁷

5 | RESULTS

We present the results in two ways. First, for each database, we report the 10 most frequently selected variables for each method. Second, we present the MSE results, the R_{OOS}^2 , and MDM test computed against the two benchmark models, and the MCS p -values are given with their ranks, with confidence levels α set to 10% and 50%, during

Models	MSE(%)	$R_{OOS}^2(AR)$	$R_{OOS}^2(DFM)$	MCS	Rank
Ridge	0.165	0.632	-0.154	0.3836	2
Lasso	0.219	0.512	-0.532	0.3778	4
EN	0.188	0.579	-0.321	0.3778	4
EN0.5	0.211	0.529	-0.478	0.3778	4
RF	0.174	0.611	-0.222	0.3836	2
DFM	0.143	0.681	0.000	1.0000	1
AR(1)	0.448	0.000	-2.139	0.2474	7

Notes: $R_{OOS}^2(AR)$ and $R_{OOS}^2(DFM)$ are the $R_{OOS}^2 = 1 - (MSE_{model}/MSE_{benchmark})$ where the benchmark model is the AR(1) model and the DFM model, respectively. * and ** Significant at the 5% and 10% level, respectively, for the MDM test of Harvey et al. (1997). MCS denotes the p -value of the statistic $T_{max.N}$ of Hansen et al. (2011) based on the MSE loss function. Rank gives model ranking position based on the $\hat{\mathcal{M}}_{90\%}^*$.

Models	MSE(%)	$R_{OOS}^2(AR)$	$R_{OOS}^2(DFM)$	MCS	Rank
Ridge	0.155	0.653	0.996	0.8528	3
Lasso	0.173	0.615	0.978	0.6909	14
EN	0.170	0.620	0.981	0.7506	13
EN0.5	0.176	0.608	0.975	0.6459	15
RF	0.173	0.614	0.978	0.6459	15
DCSIS+EN05	0.160	0.642	0.991	0.7788	5
DCSIS+EN	0.145	0.676	1.006	0.9012	2
DCSIS+Lasso	0.160	0.643	0.991	0.7788	5
DCSIS+Ridge	0.159	0.644	0.992	0.7788	5
DCSIS+RF	0.160	0.643	0.991	0.7788	5
MDCSIS+EN05	0.162	0.638	0.989	0.7788	5
MDCSIS+EN	0.145	0.677	1.006	1.000	1
MDCSIS+Lasso	0.160	0.643	0.991	0.7788	5
MDCSIS+Ridge	0.159	0.644	0.992	0.7788	5
MDCSIS+RF	0.163	0.637	0.988	0.7788	5
DFM	0.151	0.663	1.000	0.8528	3
AR(1)	0.448	0.000	0.703	0.2153	17

Notes: $R_{OOS}^2(AR)$ and $R_{OOS}^2(DFM)$ are the $R_{OOS}^2 = 1 - (MSE_{model}/MSE_{benchmark})$ where the benchmark model is the AR(1) model and the DFM model, respectively. * and ** Significant at the 5% and 10% level, respectively, for the MDM test of Harvey et al. (1997). MCS denotes the p -value of the statistic $T_{max.N}$ of Hansen et al. (2011) based on the MSE loss function. Rank gives model ranking position based on the $\hat{\mathcal{M}}_{90\%}^*$.

TABLE 6 Nowcasting performance during the full out-of-sample period - 2016 M1-2021M12 - for the small database.

TABLE 7 Nowcasting performance during the full out-of-sample period - 2016 M1-2021M12 - for the large database.

the full out-of-sample (OOS) period (2016 M1-2021M12), and also during the non-crisis (2016 M1-2019M12) and crisis (2020 M1-2021M12) periods.

5.1 | Variable selection

Table 5 displays the 10 most important variables selected by the penalized regression, with or without a screening step, for the regional and large databases. For the regional databases, the Lasso and EN methods select almost the same most important predictors, principally business surveys in the industry sector, with notably a variable on employment (EMP PdL industry BdF). When adding the national business survey, the two penalized regressions without a selection step retain only

a few regional surveys as major predictors, namely, three and one predictors for the Lasso and EN methods, respectively, whereas the regional variables are not among the most important variables for the penalized regressions with a dimensionally reduction procedure, except the BdF Business sentiment indicator in the services sector.

For the large database, we observe that only one variable is selected by all the methods with the Change in foreign orders in the industry sector provided by the BdF, suggesting that this variable could be a relevant predictor. The main difference between the regularized methods without a screening step is that the EN method selects the INSEE business surveys rather than the Lasso method, which is more focused on the BdF surveys. An interesting result is that both penalized

TABLE 8 Nowcasting performance during the full out-of-sample period - 2016 M1-2021M12 - for the regional and large databases.

Models	MSE(%)	$R_{OOS}^2(AR)$	$R_{OOS}^2(DFM)$	MCS	Rank
<i>Regional database</i>					
Ridge	0.165	0.632	-0.154	0.7623	5
Lasso	0.219	0.512	-0.532	0.2307	22
EN	0.188	0.579	-0.321	0.6817	19
EN0.5	0.211	0.529	-0.478	0.2329	21
RF	0.174	0.611	-0.222	0.7428	16
DFM	0.143	0.681	0.000	1.0000	1
AR(1)	0.448	0.000	-2.139	0.2362	20
<i>Large database</i>					
Ridge	0.155	0.653	-0.089	0.7755	3
Lasso	0.173	0.615	-0.210	0.7623	5
EN	0.170	0.620	-0.194	0.7623	5
EN0.5	0.176	0.608	-0.232	0.7428	16
RF	0.173	0.614	-0.213	0.7428	16
DCSIS+EN05	0.160	0.642	-0.122	0.7623	5
DCSIS+EN	0.145	0.676	-0.017	0.9668	2
DCSIS+Lasso	0.160	0.643	-0.121	0.7623	5
DCSIS+ridge	0.159	0.644	-0.116	0.7623	5
DCSIS+RF	0.160	0.643	-0.121	0.7623	5
MDCSIS+EN05	0.162	0.638	-0.135	0.7623	5
MDCSIS+EN	0.145	0.677	-0.014	0.9668	2
MDCSIS+Lasso	0.160	0.643	-0.120	0.7623	5
MDCSIS+ridge	0.159	0.644	-0.118	0.7623	5
MDCSIS+RF	0.163	0.637	-0.140	0.7623	5
DFM	0.151	0.663	-0.058	0.9516	4

Notes: $R_{OOS}^2(AR)$ and $R_{OOS}^2(DFM)$ are the $R_{OOS}^2 = 1 - (MSE_{model}/MSE_{benchmark})$ where the benchmark model is the AR(1) model and the DFM model estimated on the regional database, respectively. * and ** Significant at the 5% and 10% level, respectively, for the MDM test of Harvey et al. (1997). MCS denotes the p -value of the statistic $T_{max, \mathcal{M}}$ of Hansen et al. (2011) based on the MSE loss function. Rank gives model ranking position based on the $\hat{\mathcal{M}}_{90\%}^*$.

methods have the Google Trend variable at the national level as a major predictor.

The regularized methods with an initial variable-selection step select similar sentiment indicators among their most frequent predictors with the OECD CLI, the EC

ESI, and the BdF BSI. The main difference between the four penalized regressions is that the DCSIS+Lasso method selects more often BdF business surveys in the services sector whereas the PMI surveys are more frequently selected by the DCSIS+EN and MDCSIS+EN methods.

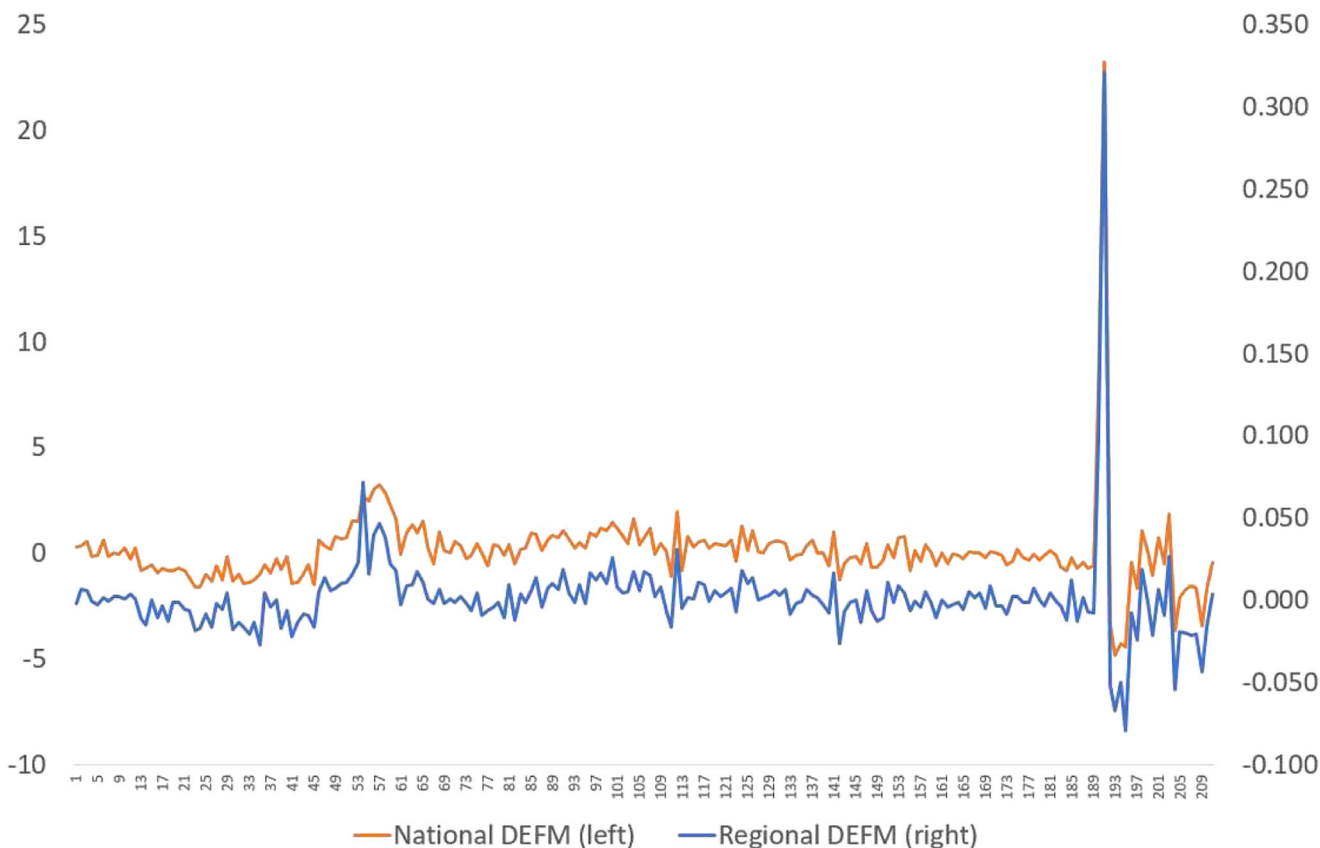


FIGURE 3 Cumulative squared prediction errors during the full out-of-sample period – 2016 M1-2021M12 – with the regional database.

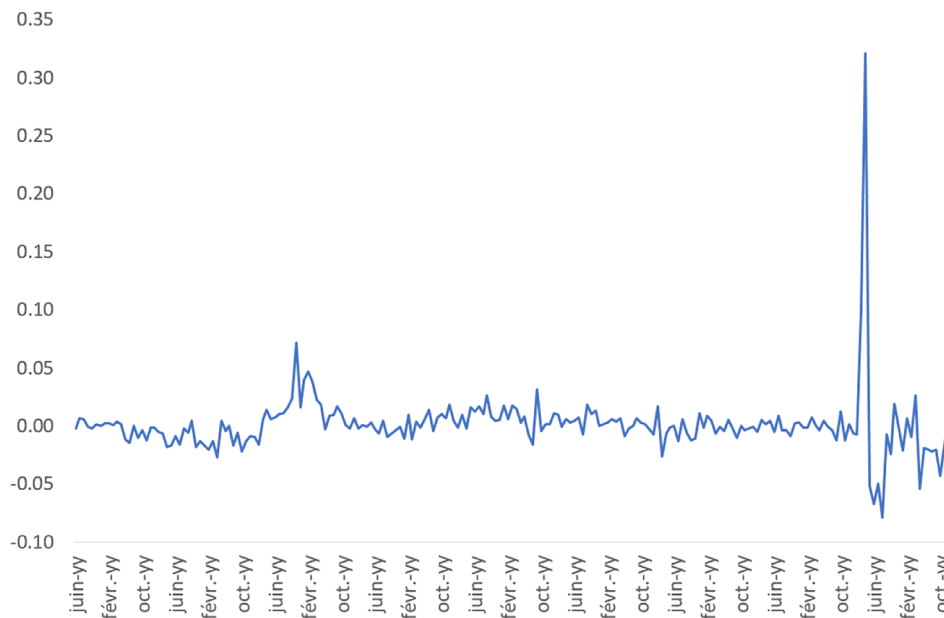


FIGURE 4 Cumulative sum of squared forecast error (CUMSFE) from the regional and large databases. Notes: cumulative sum of squared forecast error (CUMSFE) loss differential of the nowcasts provided by the DCSIS+EN model or the DFM estimated from the large database against those obtained from the EN model or the DFM estimated from the regional database.

5.2 | Nowcasting performance

Tables 6–8 present the results for the full out-of-sample period (2016 M1–2021M12). When the regional database is considered (Table 6), the best method is the DFM, producing the smallest MSE (0.142), with a large positive value of the R^2_{OOS} and a significant MDM test against the AR(1) model, and being the only model with an MCS p -value above 0.90. Penalized regressions and Random forest are also interesting alternatives since they exhibit MSE relatively close to that of the DFM (ranging between 0.165 and 0.216), with significant MDM tests against the AR(1) benchmark. Further, they are included in the set of superior models (SSM) with MCS p -values above 0.30, thus similar nowcasts with respect to the DFM. This finding is confirmed by Figure 3, which provides time-series

plots of the cumulative squared prediction error (CSPE) for the penalized models and the DFM to assess their nowcasting performance over time.

Results for the large database highlight the performance of the Elastic-Net model with a SIS-type pre-selection, producing the smallest MSEs (Table 7). The DCSIS+EN and MDCSIS+EN models outperform the AR(1) benchmark, with large positive values of R^2_{OOS} , significant MDM tests, and are the only models within the $\widehat{\mathcal{M}}_{10\%}^*$. All other models display similar nowcasting performance since they are within the $\widehat{\mathcal{M}}_{50\%}^*$, except the AR(1) model. Another interesting result is that the Lasso, Elastic-Net, and RF models combined with a screening procedure exhibit better MSE than their counterparts without pre-selection. This finding confirms those obtained by Borup and Schütte (2022) and Borup et al.

TABLE 9 Nowcasting performance during the non-crisis period - 2016 M1–2019M12 – for the regional and large database.

Models	MSE(%)	$R^2_{OOS}(AR)$	$R^2_{OOS}(DFM)$	MCS	Rank
<i>Regional database</i>					
Ridge	0.0070	0.273	0.148	0.9992	4
Lasso	0.0072	0.247	0.119	0.9953	13
EN	0.0071	0.258	0.131	0.9964	11
EN0.5	0.0071	0.261	0.134	0.9969	8
RF	0.0072	0.248	0.119	0.9953	13
DFM	0.0082	0.146	0.000	0.9586	21
AR(1)	0.0096	0.000	−0.171	0.7059	23
<i>Large database</i>					
Ridge	0.0071	−0.024	0.135	0.9969	8
Lasso	0.0070	−0.009	0.148	0.9975	5
EN	0.0069	0.000	0.156	1.0000	1
EN0.5	0.0069	0.000	0.156	1.0000	1
RF	0.0069	0.000	0.156	1.0000	1
DCSIS+EN05	0.0071	−0.031	0.130	0.9964	10
DCSIS+EN	0.0078	−0.132	0.045	0.9752	20
DCSIS+Lasso	0.0072	−0.047	0.117	0.9953	13
DCSIS+ridge	0.0071	−0.021	0.138	0.9975	5
DCSIS+RF	0.0077	−0.120	0.055	0.9829	19
MDCSIS+EN05	0.0074	−0.068	0.099	0.9932	17
MDCSIS+EN	0.0086	−0.242	−0.049	0.5569	24
MDCSIS+Lasso	0.0073	−0.051	0.113	0.9950	16
MDCSIS+ridge	0.0072	−0.042	0.120	0.9959	12
MDCSIS+RF	0.0076	−0.096	0.075	0.9921	18
DFM	0.0080	−0.153	0.027	0.9357	22

Notes: $R^2_{OOS}(AR)$ and $R^2_{OOS}(DFM)$ are the $R^2_{OOS} = 1 - (MSE_{model}/MSE_{benchmark})$ where the benchmark model is the AR(1) model and the DFM model estimated on the regional database, respectively. * and ** Significant at the 5% and 10% level, respectively, for the MDM test of Harvey et al. (1997). MCS denotes the p -value of the statistic $T_{max,\mathcal{M}}$ of Hansen et al. (2011) based on the MSE loss function. Rank gives model ranking position based on the $\widehat{\mathcal{M}}_{90\%}^*$.

Models	MSE(%)	$R_{OOS}^2(AR)$	$R_{OOS}^2(DFM)$	MCS	Rank
<i>Regional database</i>					
Ridge	0.480	0.637	-0.166	0.7181	9
Lasso	0.641	0.516	-0.558	0.1734	22
EN	0.551	0.584	-0.339	0.5940	21
EN0.5	0.618	0.533	-0.503	0.1737	23
RF	0.508	0.616	-0.235	0.7181	9
DFM	0.412	0.689	0.000	1.0000	1
AR(1)	1.324	0.000	-2.217	-	24
<i>Large database</i>					
Ridge	0.452	0.659	-0.098	0.9202	5
Lasso	0.504	0.620	-0.224	0.7181	9
EN	0.497	0.624	-0.208	0.7181	9
EN0.5	0.513	0.612	-0.247	0.7181	9
RF	0.505	0.619	-0.227	0.6985	20
DCSIS+EN05	0.466	0.648	-0.132	0.7181	9
DCSIS+EN	0.419	0.683	-0.019	0.9678	2
DCSIS+Lasso	0.465	0.649	-0.131	0.7181	9
DCSIS+ridge	0.463	0.650	-0.126	0.7181	9
DCSIS+RF	0.464	0.649	-0.128	0.7899	6
MDCSIS+EN05	0.471	0.644	-0.144	0.7181	9
MDCSIS+EN	0.417	0.685	-0.012	0.9678	2
MDCSIS+Lasso	0.465	0.649	-0.129	0.8309	7
MDCSIS+ridge	0.464	0.650	-0.127	0.7311	8
MDCSIS+RF	0.473	0.643	-0.148	0.7181	9
DFM	0.437	0.670	-0.062	0.9532	4

Notes: $R_{OOS}^2(AR)$ and $R_{OOS}^2(DFM)$ are the $R_{OOS}^2 = 1 - (MSE_{model}/MSE_{benchmark})$ where the benchmark model is the AR(1) model and the DFM model estimated on the regional database, respectively. * and ** Significant at the 5% and 10% level, respectively, for the MDM test of Harvey et al. (1997). MCS denotes the p -value of the statistic $T_{max,\mathcal{M}}$ of Hansen et al. (2011) based on the MSE loss function. Rank gives model ranking position based on the $\vec{\mathcal{M}}_{90\%}$.

(2023), showing that pre-selecting predictors improve the RF model. However, this is not the case for the Ridge regression.

Finally, Table 8 compares the nowcasting performance of the models estimated on the two databases. Results show that the SSM with MCS p -values above 0.90 includes the DFMs estimated on the regional and large databases as well as the EN models with a screening step estimated on the large database. This finding suggests that adding the national surveys to the regional database does not yield gains in predictive accuracy for the DFM, whereas it improves the nowcasting performance of the penalized regressions combined with a screening procedure.

To better see whether adding the national surveys to the regional database can improve the predictive ability,

TABLE 10 Nowcasting performance during the crisis period - 2020 M1-2021 M12 - for the regional and large database.

Figure 4 plots the cumulative sum of squared forecast error (CUMSFE) loss differential of the nowcasts provided by the EN model and the DFM estimated with the large database against those estimated with the regional database. The CUMSFE is given by

$$\text{CUMSFE}_{t,t+k} = \sum_{i=t}^{t+k} e_{i,M_1}^2 - e_{i,M_2}^2$$

where M_1 is the benchmark, in our case the EN model or the DFM, estimated with the regional database, and M_2 is the DCSIS+EN model or the DFM estimated with the large database. The given model outperforms the benchmark up to $t+k$ if the $\text{CUMSFE}_{t,t+k}$ is positive. Figure 4 shows that both models estimated with the regional and large databases did not display significant differences

until the COVID-19 crisis since their CUMSFE is very close to 0. We observe the largest gains of additional national survey data at the beginning of the COVID-19 crisis. However, during the COVID-19 crisis, both models show opposite behavior since the DCSIS+EN model exhibits positive values, suggesting that the nowcasts obtained with the large database outperform those obtained with the small database, whereas it is the opposite for the DFM for which the CUMSFE decreases.

Figures 3 and 4 have shown a strong jump associated with the COVID-19 crisis. Thus, we now analyze whether some models are more adequate during the crisis period and other models during the non-crisis period. During the non-crisis period (2016 M1–2019M12), all the models have a similar predictive ability since all have close MSE and belong to the SSM with a confidence level 10%, except the AR(1) model, for the regional and large databases (Table 9). The lowest MSEs are given by the EN and RF models estimated on the large database. Note that the two-step procedures with a prior screening approach do not seem to improve the nowcasting accuracy during the non-crisis period.

When we consider the COVID-19 crisis period (2020 M1–2021M12) the SSM based on the $\widehat{\mathcal{M}}_{10\%}^*$ contains the DCSIS+EN, MCDSIS+EN, and DFM estimated on the large database as well as the DFM estimated on the regional data (Table 10). All other models are within $\widehat{\mathcal{M}}_{30\%}^*$, while the Lasso, EN, and AR(1) models estimated on the regional database are excluded from this SSM. An interesting finding is that adding the national surveys to the regional database seems to have benefits on the nowcast accuracy for the penalized regressions. This could be explained by the fact that the national surveys have better taken into account the COVID-19 crisis than the regional surveys.

6 | CONCLUSION

This paper developed nowcasting models for the jobseekers of PdL by using penalized regressions, random forest, and dynamic factor models applied to a broad set of regional and national predictors. The results showed that when adding the national surveys to the regional database it appears that the nowcasting performance is improved for the penalized regressions whereas it is not the case for the dynamic factor model. In particular, the Elastic-Net model with the DCSIS and MDCSIS preselections display accuracy gains during the COVID-19 crisis in adding national surveys, suggesting that national surveys seem to be informative during this crisis period.

From a practitioner's viewpoint, our results indicate that the EN model with a screening step is a relevant

approach to nowcast the PdL job seekers since they give similar nowcasting accuracy to the DFM and allow for identifying the main predictors. For example, the change in foreign orders in the industry sector provided by the BdF, the OECD Composite Leading Indicator, and the BdF Business Sentiment Indicator.

This paper is focused on nowcasting the PdL's jobseeker growth rate. Further research would be to nowcast the jobseekers for the other French regions. Even the Google Trend variables did not seem to be a major predictor we could construct indicators based on Google Trends as suggested by Baker and Fradkin (2017) or Caperna et al. (2022). We can also focus on youth jobseekers for the age group 15–24 years in order to fully exploit the Google Trend variable because internet use is most likely affected by a generation bias (D'Amuri, 2009; Fondeur & Karamé, 2013). For the machine learning approaches, we used penalized regressions. We could also consider other approaches that allow nonlinear specification in future research, such as decision trees or support vector machines as suggested by Ahmad et al. (2023) and Gogas et al. (2022) for unemployment forecasting.

ACKNOWLEDGMENTS

We would like to express our gratitude to the members of LEMNA (Laboratoire d'Economie et de Management de Nantes Atlantique) for their invaluable support and resources throughout the course of this research. Special thanks are extended to Professor Olivier Darné and Dr. Amélie Charles for their guidance, expertise and insightful feedback which greatly enriched this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ENDNOTES

¹ Since 2016 there are 13 regions in continental France: Auvergne-Rhône-Alpes, Bourgogne-Franche-Comté, Bretagne, Centre-Val de Loire, Corse, Grand Est, Hauts-de-France, Île-de-France, Normandie, Nouvelle-Aquitaine, Occitanie, Pays de la Loire and Provence-Alpes-Côte d'Azur.

² *Pôle Emploi* has been formed in 2008 by merging the ANPE (National Employment Agency) and the ASSEDIC network (Association for Employment in Industry and Commerce). Since 1st January 2024 *France Travail* replaces *Pôle Emploi* in accordance with the law on full employment of 18 December 2023.

³ The other categories are: B for jobseekers working for a short period of time (up to 78 hours a month); C for jobseekers working for a long period of time (more than 78 hours a month); D for non-available jobseekers (because of learning, disease, for instance); and E for jobseekers already having a job. See Le

Barbanchon and Malherbet (2013) for an anatomy of the French labor market.

⁴ We also applied other approaches of DFMs with the dynamic PCA in the frequency domain (Forni et al. (2005), and the DFM estimated from quasi-maximum likelihood (Doz et al., 2012). The results have been lower than those obtained from the PCA model and are available from the authors upon request.

⁵ See Liu et al. (2015) for a selective survey on screening methods.

⁶ Due to the fact that the MDM test can be biased for nested models we have also applied the Clark and McCracken (2001) encompassing test (ENC-NEW). The results are similar to those with the MDM test, and are available from the authors upon request.

⁷ The MCS *p*-values are calculated through 10,000 stationary block bootstraps (Hansen et al., 2011).

REFERENCES

- Abberger, K. (2007). Qualitative business surveys and the assessment of employment - a case study for Germany. *International Journal of Forecasting*, 23, 249–258. <https://doi.org/10.1016/j.ijforecast.2006.10.002>
- Ahmad, M., Khan, Y. A., Jiang, C., Kazmi, S. J. H., & Abbas, S. Z. (2023). The impact of COVID-19 on unemployment rate: An intelligent based unemployment rate prediction in selected countries of Europe. *International Journal of Finance and Economics*, 28, 528–543. <https://doi.org/10.1002/ijfe.2434>
- Baker, S. R., & Fradkin, A. (2017). The impact of unemployment insurance on job search: Evidence from Google search data. *The Review of Economics and Statistics*, 99, 756–768. https://doi.org/10.1162/REST_a_00674
- Borup, D., Christensen, B. J., Mühlbach, N. S., & Nielsen, M. S. (2023). Targeting predictors in random forest regression. *International Journal of Forecasting*, 39, 841–868. <https://doi.org/10.1016/j.ijforecast.2022.02.010>
- Borup, D., & Schütte, E. C. M. (2022). In search of a job: Forecasting employment growth using Google trends. *Journal of Business & Economic Statistics*, 40, 186–200.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Caperna, G., Colagrossi, M., Geraci, A., & Mazzarella, G. (2022). A babel of web-searches: Googling unemployment during the pandemic. *Labour Economics*, 74, 102097. <https://doi.org/10.1016/j.labeco.2021.102097>
- Choi, H., & Varian, H. (2009). *Predicting initial claims for unemployment benefits* (Vol. 1) (pp. 1–5). Google Inc.
- Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105, 85–110. [https://doi.org/10.1016/S0304-4076\(01\)00071-9](https://doi.org/10.1016/S0304-4076(01)00071-9)
- D'Amuri, F. (2009). *Predicting unemployment in short samples with internet job search queries data*. University Library of Munich.
- D'Amuri, F., & Marcucci, J. (2017). The predictive power of google searches in forecasting US unemployment. *International Journal of Forecasting*, 33, 801–816. <https://doi.org/10.1016/j.ijforecast.2017.03.004>
- Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics*, 94(4), 1014–1024.
- Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of the Royal Statistical Society, Series B*, 70, 849–911. <https://doi.org/10.1111/j.1467-9868.2008.00674.x>
- Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, 30, 117–125. <https://doi.org/10.1016/j.econmod.2012.07.017>
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100(471), 830–840.
- Gogas, P., Papadimitriou, T., & Sofianos, E. (2022). Forecasting unemployment in the euro area with machine learning. *Journal of Forecasting*, 41, 551–566. <https://doi.org/10.1002/for.2824>
- Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37, 920–964. <https://doi.org/10.1002/jae.2910>
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). Model confidence sets for forecasting models. *Econometrica*, 79, 453–497. <https://doi.org/10.3982/ECTA5771>
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291.
- Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178, 352–367. <https://doi.org/10.1016/j.jeconom.2013.08.033>
- Kim, H. H., & Swanson, N. R. (2018). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34, 339–354. <https://doi.org/10.1016/j.ijforecast.2016.02.012>
- Kotchoni, R., Leroux, M., & Stevanovic, D. (2019). Macroeconomic forecast accuracy in a data-rich environment. *Journal of Applied Econometrics*, 34, 1050–1072. <https://doi.org/10.1002/jae.2725>
- Le Barbanchon, T., & Malherbet, F. (2013). An anatomy of the French labour market: Country case study on labour market segmentation. Employment Working Paper No. 142, International Labour Organization.
- Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30, 996–1015. <https://doi.org/10.1016/j.ijforecast.2014.03.016>
- Li, R., Zhong, W., & Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107, 1129–1139. <https://doi.org/10.1080/01621459.2012.695654>
- Liu, J., Zhong, W., & Li, R. (2015). A selective overview of feature screening for ultrahigh-dimensional data. *SCIENCE CHINA Mathematics*, 58, 1–22. <https://doi.org/10.1007/s11425-015-5062-9>
- Nagao, S., Takeda, F., & Tanaka, R. (2019). Nowcasting of the US unemployment rate using Google trends. *Finance Research Letters*, 30, 103–109. <https://doi.org/10.1016/j.frl.2019.04.005>
- Niesert, R. F., Oorschot, J. A., Veldhuisen, C. P., Brons, K., & Lange, R.-J. (2020). Can Google search data help predict macroeconomic series? *International Journal of Forecasting*, 36, 1163–1172. <https://doi.org/10.1016/j.ijforecast.2018.12.006>

- Silverstovs, B. (2013). Do business tendency surveys help in forecasting employment? A real-time evidence for Switzerland. *Journal of Business Cycle Measurement and Analysis*, 2013, 129–151. <https://doi.org/10.1787/jbcm-2013-5k4bxlxjkd32>
- Simionescu, M., & Cifuentes-Faura, J. (2022). Forecasting national and regional youth unemployment in Spain using Google trends. *Social Indicators Research*, 164, 1187–1216. <https://doi.org/10.1007/s11205-022-02984-9>
- Soybilgen, B., & Yazga, E. (2018). Evaluating nowcasts of bridge equations with advanced combination schemes for the Turkish unemployment rate. *Economic Modelling*, 72, 99–108. <https://doi.org/10.1016/j.econmod.2018.01.009>
- Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162. <https://doi.org/10.1198/073500102317351921>
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35, 2769–2794. <https://doi.org/10.1214/009053607000000505>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429. <https://doi.org/10.1198/016214506000000735>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67, 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>

AUTHOR BIOGRAPHIES

Clément Cariou has a PhD in economics from IAE Nantes and works in particular on time series forecasting and labour market issues.

Amélie Charles is a professor in finance at the Audencia Business School. Her main research interests include topics in forecasting, volatility, financial modelling and effects of outliers.

Olivier Darné is a professor in economics at the Nantes University (LEMNA). His main research interests include topics in time series modelling, nowcasting, volatility and macroeconometrics.

How to cite this article: Cariou, C., Charles, A., & Darné, O. (2024). Are national or regional surveys useful for nowcasting regional jobseekers? The case of the French region of Pays-de-la-Loire. *Journal of Forecasting*, 43(6), 2341–2357. <https://doi.org/10.1002/for.3125>