Cyprus Economic Policy Review, Vol. 16, No. 1, pp 11-36 (2022) 1450-4561

Forecasting economic activity using preselected predictors: the case of Cyprus

Christiana Anaxagorou and Nicoletta Pashourtidou*

Economics Research Centre, University of Cyprus

Abstract

This paper applies hard and soft thresholding techniques to a large dataset of domestic and foreign series in order to preselect informative predictors for forecasting economic activity in Cyprus. The variables of interest in the forecasting exercise consist of GDP and the production-side components of GDP, expressed in growth rates. The subsets of selected predictors are allowed to differ across the variables of interest and over the forecast horizon, thus accommodating idiosyncratic features of economic sectors. The sets of selected predictors contain a higher proportion of domestic as opposed to foreign predictors for the one-quarter horizon, while the opposite occurs for longer horizons. Furthermore, in the case of GDP all thresholding techniques result in selecting high proportions of business and consumer survey indicators for all horizons. The forecasting performance depends on the forecast horizon and, most importantly, on whether the subsets of chosen predictors remain constant or change over time. The thresholding technique employed is not found to substantially affect the forecasting performance. Selecting predictors prior to forecasting GDP growth, leads to lower forecast errors vis-à-vis a simple univariate benchmark, as well as compared to exploiting the full dataset of predictors for forecasting. The gains from preselecting predictors are higher during a crisis period than in normal times, especially for short horizons, while preselection in normal times benefits forecast accuracy for longer horizons. Predictor preselection is found to improve the forecasting performance in the case of some production-side components, particularly the gross value added in the sectors of trade and construction, and net taxes.

Keywords: Forecasting, factor models, predictor selection

1. Introduction

Effective economic decisions by policy makers or private agents are dependent on the available information. Economic agents and policy makers, however, are likely to have imperfect information on various indicators of economic performance at the aggregate and sectoral levels, because of delays in the publication of official data, particularly national accounts, highlighting the importance of forecasting economic activity in economic and financial decision-making. The quantity of available data has increased rapidly in recent years, mainly because of technological improvements and the rise of big data relating to economic and financial indicators. The availability of many

* Corresponding Author Address: Economics Research Centre, University of Cyprus, P.O. Box 20537, 1678, Nicosia, Cyprus. Email: n.pashourtidou@ucy.ac.cy.

candidate predictors has given rise to forecasting methods that exploit the richness of information without proportionally increasing the number of estimated parameters. Such methods, which include forecast combinations (or forecast pooling), dynamic factor models and Bayesian model averaging, have been widely applied and found to improve upon the forecasting performance of univariate benchmarks for many macroeconomic variables (see Stock & Watson, 2006 for a survey).

The various indicators that can be employed to forecast activity developments are released according to different publication schedules and with different frequencies, resulting in ragged-edge data and mixed frequency problems that have been extensively studied in the literature (see e.g. Stock & Watson, 2002; Kitchen & Monaco, 2003; Baffigi, Golinelli & Parigi, 2004; Giannone, Reichlin & Small, 2008; Banbura & Runstler, 2011; Foroni & Marcellino, 2014). Typically, an economic activity variable in quarterly frequency such as GDP, consumption, or sectoral value added, is forecast using indicators published on a monthly basis, stressing the need for timely available monthly data. While the plethora of available data, along with dimension reduction methods, can assist practitioners and policy makers in their efforts to forecast economic activity, the use of many noisy indicators (usually of higher publication frequency than the variable to be forecast) may not necessarily improve the forecasting performance. There is a growing body of literature that investigates the effects of indicator preselection from large datasets on the accuracy of forecasts for macroeconomic variables (see e.g. Boivin & Ng, 2006; Bai & Ng, 2008; Kim & Swanson, 2014, 2018; Jokubaitis, Celov & Leipus, 2021).

Prior to forecasting, indicators can be selected from a large dataset by applying "hard" thresholding rules, i.e. a statistical significance test, or "soft" thresholding methods, i.e. optimisation algorithms that carry out variable selection and shrinkage at the same time, providing also a ranking of the predictors (see e.g. Bai & Ng, 2008). The information in selected predictors is usually summarized through a small number of estimated common factors, which are subsequently used in forecasting models. The simulations in Boivin and Ng (2006) show that factors extracted from larger panels of series could lead to higher forecast errors compared to factors estimated from smaller datasets; this occurs when the extended panel includes series with cross-correlated idiosyncratic errors and/or higher error variances than those in the smaller panel. The Monte Carlo study in Boivin and Ng (2006) also demonstrates that the forecasting power of factors estimated from a given panel depends on their dominance in the panel (i.e. the size of factor loadings) and, therefore, on how informative the dataset is about the factors. A factor that is dominant in a smaller dataset can lead to inferior forecasts, if it is dominated by another factor in a larger dataset. The empirical analysis in Boivin and Ng (2006) that includes many economic variables reveals that factors extracted from the full set of 147 series do not yield superior forecasts vis-à-vis factors estimated from smaller sets of prescreened indicators.

Empirical studies on forecasting GDP growth in the euro area as a whole and in large euro area economies, have documented gains associated with applying preselection techniques to large datasets vis-à-vis utilising the full set of predictors (see e.g. Schumacher, 2010; Caggiano, Kapetanios & Labhard, 2011; Bulligan, Marcellino & Venditti, 2015; Girardi, Golinelli & Pappalardo, 2017). Moreover, the empirical results of Caggiano, Kapetanios and Labhard (2011) highlight the benefits of combinations of factor-based forecasts obtained from prescreened datasets.

Empirical evidence, particularly for the euro area, provides the motivation for considering subsets of predictors drawn from large datasets in forecasting economic activity in Cyprus. Previous work using data for Cyprus has shown that GDP and sectoral growth forecasts computed from a set of preselected predictors, using a simple hard thresholding method (i.e. significant correlations with dependent variables), are at least as accurate as the forecasts obtained using the full set of predictors (Pashourtidou, Papamichael & Karagiannakis, 2018). This paper builds on previous work on the performance of methods for forecasting activity growth in Cyprus by applying hard and soft thresholding techniques to a large dataset of domestic and foreign series in order to preselect informative predictors. The variables of interest in the forecasting exercise are GDP and its production-side components, expressed in growth rates. The subsets of selected predictors are allowed to differ across the variables of interest and over the forecast horizon, thus addressing the idiosyncratic features of sectors. Our results show that the forecasting performance depends on the forecast horizon and on whether the subsets of chosen predictors remain constant or change over time. The forecasting performance is not found to vary considerably with the thresholding technique employed for selecting the predictors. Selecting a subset of predictors before forecasting GDP growth, leads to lower forecast errors vis-à-vis a simple univariate benchmark, as well as compared to exploiting the full dataset of predictors for forecasting; the gains from preselecting predictors are higher during a crisis period than in normal times. Predictor preselection is found to improve the forecasting performance in the case of some production-side components, particularly the gross value added in the sectors of trade and construction, and net taxes.

The structure of the paper is as follows. Section 2 presents the forecasting model when a large number of predictors is used and gives an overview of predictor selection methods. Section 3 briefly describes the dataset and details the thresholding techniques and schemes employed for preselecting predictors from the full dataset. Section 4 describes the forecasting exercise and presents the results; it also discusses the forecasting performance during crisis and normal periods. Section 5 concludes.

2. Forecasting with preselected predictions

We are interested in forecasting a variable, y_t, using data on a large number of predictors, $X_t = \left(X_{1,t}, \dots, X_{N,t}\right)'$, and $t = 1, \dots, T$. In many empirical applications the number of available predictors, N, is larger than the number of available time periods, T. The factor approach is used to solve the dimensionality problem and obtain feasible forecasts for y_t, while utilising the information in X_t. In the factor approach, the h-step ahead forecast, y_{t+h}^h, for y_t is given by the model

$$y_{t+h}^h = \alpha_h + \sum_{j=0}^{p} \beta_{h,j}' F_{t-j} + \sum_{m=0}^{q} \gamma_{h,m} y_{t-m} + \varepsilon_{t+h}, \tag{1}$$

and the series of predictors have a factor representation

$$X_{i,t} = \lambda_i' F_t + e_{i,t}, \qquad i = 1, \dots, N \tag{2}$$

where F_t is an $r \times 1$ vector of factors, i.e. a component common to all predictors, λ_i is a vector of factor loadings for series i and e_{it} is the idiosyncratic error. As the vector of factors is unobserved, an estimate, $\hat{F}_t$, is obtained using principal components.

Equation (1) is then estimated with data up to period $T-h$ and by replacing F_t with $\hat{F}_t$.[1] The h-step ahead forecast is computed as

$$\hat{y}^h_{T+h} = \hat{\alpha}_h + \sum_{j=0}^{p} \hat{\beta}'_{h,j}\hat{F}_{T-j} + \sum_{m=0}^{q} \hat{\gamma}_{h,m} y_{T-m}. \quad (3)$$

The number of factors is typically estimated using information criteria based solely on the variation in the panel of predictors, X_t, without involving the variable to be forecast. However, estimated factors that capture a high proportion of the variation in the panel may not necessarily have a strong predictive content for the variable of interest. An issue investigated in the literature is how the quantity of information, i.e. the size of N, affects the accuracy of the forecasts in (3) through the quality of information conveyed by the estimated factors in $\hat{F}_t$. The effects of the number of predictors (N) on the forecasting performance of models similar to (1) in terms of mean squared forecast error have been widely investigated in the literature. Applications cover various large countries and a range of variables. For example, Bai and Ng (2008) focus on US inflation; Caggiano, Kapetanios and Labhard (2011) look at GDP growth for euro area countries and the UK; Kim and Swanson, (2014, 2018) consider key macroeconomic variables for the US; Bulligan, Marcellino and Venditti (2015) examine GDP growth for Italy with reference to other major euro area member states and the euro area as a whole; Panagiotelis et al. (2019) study main macroeconomic indicators for Australia. The above empirical applications use hard and/or soft thresholding techniques for selecting the predictors employed in the estimation of factors, which are subsequently included in forecasting equations.

Under a hard thresholding method, a predictor is chosen based on its statistical significance for the variable of interest; for example, a test for the significance of the correlation coefficient between a predictor and the variable of interest can be applied. Hard thresholding methods ignore information in other candidate predictors and may result in selecting collinear predictors, i.e. indicators whose information content is very similar. Another limitation of hard thresholding is that the set of selected predictors depends on the choice of the test or the decision rules used.

Soft thresholding methods are algorithms that select predictors and perform shrinkage, i.e. reduce the size of the estimates associated with predictors that are not important for the variable of interest. A widely used method, known as the Least Absolute Shrinkage Selection Operator (LASSO), performs a penalised regression on all available predictors with a penalty that results in some coefficient estimates being exactly equal to zero. In other words, LASSO selects a subset of predictors by shrinking to zero the coefficient estimates of uninformative regressors and therefore excluding the latter from the active set of variables. Another popular model selection algorithm is the forward selection method. The algorithm starts by adding to the null model the predictor with the maximum correlation with the variable of interest; at each step the predictor with the maximum correlation with the residual from the previous step is selected, until the model cannot be improved further (based on a statistical criterion) by including additional regressors. A shortcoming of the forward selection algorithm is that it excludes useful predictors that are correlated with the regressors selected in previous steps. A more general method, the Least Angle Regression (LARS) is a less aggressive selection algorithm than the forward selection method (see Efron et al.,

[1]Bai and Ng (2006) show that estimating (1) by least squares when F_t is replaced by the principal component estimate, $\hat{F}_t$, leads to consistent and asymptotically normal parameter estimates and forecasts; hence, the estimated factors in $\hat{F}_t$ can be treated as observed variables.

2004). More specifically, LASSO and a more cautious variant of the forward selection algorithm, known as the forward stagewise regression, can be implemented through modifications of the LARS method. Bai and Ng (2008) list the advantages of LARS. For example, LARS utilises the information in all candidate predictors and ranks them in order of importance with respect to the variable of interest; it excludes highly correlated predictors without being as aggressive as the forward selection method; and it is fast to compute.

In the empirical analysis that follows, we apply hard and soft thresholding methods to a large dataset of macroeconomic and financial indicators in order to select a subset of predictors that are informative for economic activity growth in Cyprus.

3. Data and predictor selection

As we focus on forecasting economic activity at the aggregate and sectoral levels, the variables of interest are GDP and the production-side components of GDP obtained from quarterly national accounts.[2] In the analysis that follows, the production-side components of GDP are given by the Gross Value Added (GVA) in 10 sectors of economic activity and net taxes on products, as shown in Table A1. The predictors consist of 200 indicators that cover domestic and external economic and financial conditions, and are published well before the release of the quarterly national accounts. The dataset includes indicators for real activity and labour market (e.g. volume indices of retail trade and manufacturing, building permits, vehicle registrations, unemployment), financial series (e.g. stock market indices, interest rates, exchange rates, loans), price indices and international commodity prices, as well as series from business and consumer surveys in Cyprus and the EU; the full list of indicators is given in Table A2.[3] Over 90% of indicators in the dataset are published on a monthly basis, while a small number of quarterly series with forward-looking properties are also included.[4] Table 1 summarises the composition of the dataset; 60% of variables are domestic indicators and the remaining 40% represents foreign or international series. The two largest groups of indicators consist of survey and financial indicators, which are usually available at the end of the reference month. The dataset spans the period from 1995Q1 to 2019Q4.

[2] National accounts data were obtained from the Statistical Service of Cyprus; the seasonally adjusted chain-linked volume measures are used.

[3] The data were obtained from various local and international sources. Local sources include the Statistical Service of Cyprus, the Central Bank of Cyprus, the Cyprus Stock Exchange, the Department of Lands and Surveys, and the Department of Registrar of Companies and Intellectual Property. The sources of foreign/international variables include: Eurostat, the European Central Bank, the European Commission, Datastream and Global Financial Data. All data are seasonally adjusted and transformed into stationary series by differencing the levels of the series, or the logarithm of the series levels.

[4] Quarterly series with forward-looking properties relate to information from the Survey of Professional Forecasters (European Central Bank) and quarterly questions in Business and Consumer Surveys in Cyprus (European Commission).

TABLE 1
Composition of the dataset

Predictors	Number of variables (percentage)
Total	200 (100%)
Domestic	117 (60%)
Foreign/international	83 (40%)
By category	
Real activity and labour market	45 (23%)
Financial	66 (33%)
Prices	18 (9%)
Business and consumer surveys	71 (36%)

In the empirical analysis, we employ hard and soft thresholding algorithms along the lines of other studies (e.g. Bai & Ng, 2008; Bulligan, Marcellino & Venditti, 2015; Girardi, Golinelli & Pappalardo, 2017) to select indicators from the panel of 200 series; the smaller set of selected indicators is then used for forecasting the growth rate of GDP and its production-side components. We apply a hard thresholding algorithm based on an F-test, as well as the LARS and LASSO methods.

As a hard thresholding rule, we consider the statistical significance of a candidate predictor after controlling for lagged values of the dependent variable. More specifically, for each candidate predictor, $X_{i,t}$, a regression of the variable of interest, y_t^h, on Z_{t-h} and $X_{i,t-h}$ is run, where Z_{t-h} is a vector of deterministic terms and lags of y_t, and $X_{i,t-h}$ is a vector of lags and monthly leads (if available) of the candidate predictor.[5] Next, the p-value, $pv_{[i]}$, corresponding to the F-statistic for testing the statistical significance of $X_{i,t-h}$, is obtained. The p-values from the N regressions are ranked in ascending order. Predictor i is selected, if $pv_{[i]}$ does not exceed the significance level of 5%.

The soft thresholding methods use optimisation techniques to rank and select indicators. The LASSO method solves the following constrained minimisation problem

$$\min_b RSS + \lambda \sum_{i=1}^{M} |\xi_i|, \tag{4}$$

where RSS is the residual sum of squares from the regression of the variable of interest on all available regressors, i.e. the N candidate indicators and predetermined variables, with regression coefficients ξ_i. Parameter λ ($\lambda \geq 0$) determines the degree of shrinkage, i.e. the extent to which coefficients are estimated as zero and therefore the corresponding indicators are not selected as relevant for the variable of interest. Thus, the LASSO regression starts from a large number of regressors and through the minimisation problem in (4), which takes into account cross-correlations among regressors, sets to zero the coefficients of regressors that are not informative for the variable of interest. The LARS method is a more general selection algorithm than LASSO, and is computationally fast (Efron et al., 2004; Bai & Ng, 2008). The LARS method is based on the idea of updating the projection of the variable of interest on a set of $k-1$ predictors with information from a newly added predictor in the active

[5] The deterministic terms include a constant and, if needed, dummy variables; the number of lags in the regression is determined by information criteria (Bayesian and Akaike).

set. The LARS method starts with all coefficients set to zero, picks the predictor with the highest correlation with the variable of interest and computes the projection; next, the algorithm searches for the predictor that is the most correlated with the current residual to update the projection. LARS then continues in an equiangular direction between the two selected variables until another predictor is added in the most correlated set and so on. After k steps there are K variables in the active set, i.e. there are K predictors with non-zero coefficients and the remaining variables have zero coefficients. The algorithm returns a set of indices that show the order according to which each variable joined the active set.

The three predictor selection methods described above are applied using recursive estimation samples, starting from the first sample of T_0 observations, performing estimation and indicator selection before increasing the sample size by one observation (i.e. quarter) in the next iteration; thus, indicators are selected anew in each iteration.[6] Moreover, the three selection methods are applied separately for each variable of interest and over the different forecast horizons. The selection methods are used to rank the predictors in the dataset according to their information content with respect to GDP and each of the production-side components.

We consider two alternative schemes for isolating the informative predictors ordered through hard and soft thresholding: (a) selection in each estimation iteration and therefore every quarter (t), i.e. for each estimation sample, and (b) selection across all iterations, i.e. over all estimation samples.[7] In scheme (a), an indicator is selected, if it is ranked among the first 20 predictors with the lowest p-values under hard thresholding, or, is ordered among the best 20 predictors according to the LASSO or LARS algorithm; in this way, the size of the selected set is kept constant over sample iterations.[8] In scheme (b), an indicator is picked if the criteria described in scheme (a) hold, and its frequency of being selected, by each method, across all samples is ranked among the 20 highest. Thus, scheme (b) isolates the best 20 predictors that are selected most often across all iterations; the lowest selection frequency is around 30%.[9] Both schemes result in a subset of 20 selected indicators for each thresholding method applied; in scheme (a) the selected predictors may differ in each estimation sample, while in scheme (b) the indicators chosen are the same for all estimation samples. Studies employing European data for forecasting growth find that the performance is favoured when the number of selected predictors is rather small (e.g. Caggiano, Kapetanios & Labhard 2011; Girardi, Golinelli & Pappalardo, 2017). The empirical analysis of Caggiano, Kapetanios and Labhard (2011) for large euro area countries and the UK shows that the highest forecasting gains are achieved when factors are estimated from quite small panels of preselected predictors, with sizes ranging from 12 (UK) to 22 (Italy) indicators.

The nature of the selected predictors across all sample iterations is outlined in Table 2 in the case of GDP. Table 2 shows the distribution of the 20 most frequently selected predictors over different types of variables. For the one-quarter horizon, the LASSO

[6] T_0 equals 24.

[7] Scheme (a) follows Bai and Ng (2008) and scheme (b) resembles the selection strategy in Bulligan, Marcellino and Venditti (2015), and Girardi, Golinelli and Pappalardo (2017).

[8] The maximum number of selected indicators by LASSO and LARS cannot exceed the sample size, if the sample size is smaller than the number of candidate predictors, as in the empirical application.

[9] The lowest frequency for which an indicator is selected under scheme (b) varies somewhat across the variables of interest; the cut-off frequencies range from 17% for the GVA component in real estate activities to 48% for the GVA component in financial and insurance activities (see also Table 2 and Table A3).

and LARS methods pick more domestic rather than foreign and international predictors; the variables chosen by the hard thresholding method are evenly divided between the domestic and foreign groups. For longer horizons all methods select a higher proportion of foreign as opposed to domestic predictors. Survey indicators are included in the sets of selected predictors by all methods at high proportions (over 30%) for all horizons. For the very short horizon, real activity and labour market variables are frequently picked, particularly by soft thresholding methods; as the horizon increases all methods tend to select more financial indicators and, to a smaller extent, price variables, and fewer real activity and labour market predictors. Furthermore, the three selection methods show a high degree of commonality with respect to the chosen indicators, which tends to increase with the forecast horizon. For the one-quarter horizon, about 40% of predictors picked by hard thresholding are the same as those selected by the soft thresholding methods, while this percentage exceeds 65% for the longest horizon. The LASSO and LARS methods result in subsets that have over 70% of the predictors in common.

The selection of more domestic as opposed to foreign indicators for the one-quarter horizon, especially by soft thresholding methods, is found for about half of the production-side components (Table A3). For a number of components, foreign variables tend to dominate the set of selected predictors; these are components that relate to sectors of economic activity with dependency on external demand, for example transport and hospitality, professional services, and entertainment and recreation. For the majority of components, the sets of selected predictors consist mostly of financial indicators, followed by survey and real activity variables, especially for shorter horizons. Overall, there are no striking differences among the various selection methods applied; the three methods pick indicators from the different categories with similar proportions.

The last row of Table 2 gives the lowest frequency with which an indicator should be picked across sample iterations by a given method, in order to be included in the set of selected predictors under scheme (b). Across all sample iterations, the 20 best indicators are selected more frequently when the hard thresholding method is used; thus, for a given cut-off frequency the hard thresholding rule tends to select more indicators than LASSO and LARS. The cut-off frequencies do not considerably vary over the forecast horizon. Similar patterns are observed for the GDP components (Table A3). For some components, particularly those for which relevant predictors are not readily available (e.g. agriculture, real estate activities), the cut-off frequencies associated with LASSO and LARS are just below 20%, but for most of the components and horizons the cut-off frequencies are around or higher than 30%.

TABLE 2

Percentages of selected predictors by type, and cut-off frequencies, GDP

Horizon (quarters)	1			4			8		
Predictor selection method	F-test	LASSO	LARS	F-test	LASSO	LARS	F-test	LASSO	LARS
Types of predictors									
Domestic	50	75	68	27	35	33	19	38	30
Foreign/international	50	25	32	73	65	67	81	62	70
By category									
Real activity and labour market	32	50	41	5	25	24	5	10	5
Financial	18	15	18	41	20	24	43	38	45
Prices	0	0	0	14	20	19	14	14	15
Business and consumer surveys	50	35	41	41	35	33	38	38	35
Cut-off frequency (%)	38	33	35	41	33	30	42	29	32

Notes: The percentages and frequencies are computed across all sample iterations and relate to scheme (b). The cut-off frequency shows the lowest frequency with which an indicator should be chosen across all sample iterations (i.e. all quarters) by a given method in order to be included in the set of selected predictors.

4. Forecasting

4.1 Forecasting exercise and overall performance

The panels of predictors selected by hard (F-test) and soft (LASSO and LARS) thresholding methods are used to extract factors and compute forecasts through the model described by equations (1) – (3). As the database includes a large number of monthly indicators, many of which are released at the end of the reference month, (e.g. financial and survey data), we can exploit information from leading indicators that falls within the forecast horizon. Thus, we extend the forecasting equation to incorporate the available forward-looking information in terms of leads, i.e.

$$\hat{y}_{T+h}^{h} = \hat{\alpha}_h + \sum_{j=0}^{p} \hat{\beta}_{h,j}' \hat{F}_{T-j} + \sum_{m=0}^{q} \hat{\gamma}_{h,m} y_{T-m} + \hat{\delta}_h' \hat{F}_{T+1}^{L}, \qquad (5)$$

where $\hat{F}_{t+1}^{L}$ is an $l \times 1$ vector of factors extracted from the panel of selected indicators with available data for some or all of the months in quarter $t+1$. $\hat{F}_{t+1}^{L}$ is estimated by principal components. In the empirical application, $\hat{y}_{t+h}^{h}$ is the annualised growth rate defined as

$$\hat{y}_{t+h}^{h} = \frac{400}{h} \ln \left(\frac{Y_{t+h}}{Y_t}\right),$$

where Y_t denotes the level of GDP or the level of a production-side component.

We employ different specifications of the factor-augmented forecasting equation (5), depending on the choice of the number of factors, r, lags, p and q, and factor leads, l. The number of factors and the lag length of $\hat{F}_t$ and y_t are chosen by the Akaike Information Criterion (AIC). An alternative specification of the forecasting equation is obtained by applying the Bayesian Information Criterion (BIC) for determining the number of factors and lags in (5).[10] Following Caggiano, Kapetanios and Labhard (2011) the number of factors is determined by AIC or BIC, thus, taking into account the variable of interest; this is in contrast to the use of other information criteria that choose the number of factors that best summarises the panel of predictors, without reference to the variable to be forecast. Caggiano, Kapetanios and Labhard (2011) argue in favour of using information criteria to select the number of factors within the forecasting model, as opposed to criteria that rely solely on the variation of predictors in the panel, since not all factors chosen by the latter criteria are necessarily informative for forecasting the variable of interest. Moreover, in each of the AIC and BIC specifications the number of factor leads is set to $l = 1$ or $l = 4$, resulting in four different forecasting equations in total.

The pseudo out-of-sample forecasting exercise uses data over the period 1995Q1 – 2019Q4; the first estimation sample consists of the first 24 observations and the sample is extended by one quarter in each iteration. The factors and forecasting equations are estimated anew in each iteration, using the panels of predictors selected according to scheme (a) and scheme (b). Forecasts for the growth rate of GDP and the growth rates of the production-side components are computed for $h = 1, \ldots, 8$, using a pseudo out-of-sample set-up that mirrors the availability of monthly indicators in real time; the forecasts are evaluated over the period 2002Q3 – 2019Q4.

The forecasting equation (5) that includes factors, as well as autoregressive terms of y_t, is used to compute the factor-augmented autoregressive (FAR) forecasts. The FAR forecasts from the four specifications are combined using two alternative forecast combinations: (i) the mean, and (ii) the discounted mean squared forecast error (DMSFE), with a discounting factor set to 0.90 (see e.g. Stock & Watson, 2004, 2006). Furthermore, the resulting combination forecasts for the 11 components of GDP are aggregated to compute bottom-up GDP growth forecasts.[11]

For comparison purposes, we compute the mean and DMSFE combinations of forecasts from Autoregressive Distributed Lag (ADL) models, using as predictors, one at a time, indicators selected through schemes (a) and (b). More specifically, the ADL forecasting equation takes the form,

$$\hat{y}_{T+h}^{h} = \hat{a}_h + \sum_{j=0}^{p} \hat{b}_{h,j} X_{s,T-j} + \sum_{m=0}^{q} \hat{c}_{h,m} y_{T-m} + \hat{d}_h X_{s,T+1}^{L}, \tag{6}$$

where $X_{s,t}$ is a selected predictor according to scheme (a) or (b), and $X_{s,t+1}^{L}$ denotes the lead, i.e. data on predictor $X_{s,t}$ available within the forecast horizon. The lag lengths of $X_{s,t}$ and y_t are chosen by AIC or BIC as in the factor-augmented equation (5).

Table 3 presents the root mean squared forecast error (RMSFE) of combinations of FAR and ADL model forecasts for GDP growth relative to that of the benchmark AR(1) model. The table also shows the relative RMSFE for GDP growth bottom-up forecasts, obtained by aggregating combinations of FAR and ADL component forecasts. The results are given for the three selection methods employed, i.e. hard thresholding (F-

[10] AIC or BIC can choose up to four factors; the upper bound for the number of factors was guided by information criteria that summarise the dataset (e.g. Bai & Ng, 2002) and sample size limitations. Furthermore, AIC or BIC can select up to two and four lags for $\hat{F}_t$ and y_t, respectively. The results for the case where AIC or BIC can select up to four lags for both $\hat{F}_t$ and y_t, when the sample size allows it, show some deterioration in the forecasting performance.

[11] Bottom-up GDP growth forecasts are computed by aggregating the growth contributions of the components (for more details, see Eurostat, 2013, Ch. 6).

test), LASSO and LARS under schemes (a) and (b). For comparison purposes, the column labelled "All predictors" presents the relative RMSFE when no selection method is applied, i.e. all predictors in the dataset are used to extract the factors included in the FAR models, and all indicators, one at a time, are included in the ADL models.

TABLE 3

Relative RMSFE, GDP

Horizon (quarters)	1				4				8			
Predictor selection method	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors
Benchmark AR(1): RMSFE	1.01				3.91				4.94			
Scheme (a): each *t*												
FAR - mean	**0.83**	0.89	0.98	0.91	0.76	**0.69**	**0.69**	**0.80**	1.01	1.04	1.02	0.99
FAR - DMSFE	**0.82**	0.89	0.99	0.90	0.76	**0.68**	**0.69**	**0.80**	1.00	1.04	1.02	0.99
FAR - mean, bottom-up	1.06	0.92	1.05	**0.86**	**0.81**	**0.74**	0.75	**0.86**	1.19	1.29	1.30	1.40
FAR - DMSFE, bottom-up	1.06	0.91	1.09	**0.85**	**0.80**	**0.74**	0.75	**0.86**	1.19	1.29	1.30	1.39
ADL - mean	**0.83**	**0.81**	**0.82**	**0.87**	0.88	**0.82**	**0.84**	**0.93**	1.00	0.97	**0.96**	1.02
ADL - DMSFE	**0.76**	**0.72**	**0.75**	**0.80**	0.87	0.82	**0.86**	**0.93**	1.02	1.08	0.98	1.01
ADL - mean, bottom-up	0.90	**0.91**	**0.91**	**0.94**	1.10	**0.89**	**0.90**	0.95	1.08	1.11	1.11	1.11
ADL - DMSFE, bottom-up	0.91	**0.90**	**0.85**	**0.92**	1.11	**0.89**	0.88	**0.95**	1.01	1.07	1.12	1.11
Scheme (b): all *t*												
FAR - mean	**0.82**	**0.71**	**0.75**	0.91	**0.56**	**0.56**	**0.49**	**0.80**	**0.85**	**0.92**	**0.90**	0.99
FAR - DMSFE	**0.82**	**0.71**	**0.75**	0.90	**0.55**	**0.56**	**0.49**	**0.80**	**0.85**	**0.92**	**0.90**	0.99
FAR - mean, bottom-up	**0.80**	**0.81**	**0.77**	**0.86**	**0.74**	**0.53**	**0.51**	**0.86**	1.28	1.04	0.95	1.40
FAR - DMSFE, bottom-up	**0.80**	**0.80**	**0.75**	**0.85**	**0.74**	**0.53**	**0.51**	**0.86**	1.29	1.04	0.93	1.39
ADL - mean	**0.77**	**0.80**	**0.80**	**0.87**	**0.75**	**0.81**	**0.80**	**0.93**	**0.86**	**0.90**	0.90	1.02
ADL - DMSFE	**0.72**	**0.72**	**0.72**	**0.80**	**0.77**	**0.80**	**0.79**	**0.93**	**0.84**	**0.89**	**0.88**	1.01
ADL - mean, bottom-up	**0.86**	**0.88**	**0.87**	**0.94**	**0.83**	**0.85**	**0.84**	0.95	0.95	1.02	0.98	1.11
ADL - DMSFE, bottom-up	**0.83**	**0.84**	**0.82**	**0.92**	**0.81**	**0.82**	**0.82**	**0.95**	0.92	1.01	0.96	1.11

Notes: The table entries show the RMSFE of each forecast method relative to that of the AR(1) model; the pseudo out-of-sample forecasts are computed using subsets of predictors selected by hard (F-test) or soft (LASSO, LARS) thresholding methods, or the full set of predictors in the dataset. Entries in bold indicate statistical significance at 10% level for the modified Diebold-Mariano test of equal forecast accuracy (Diebold & Mariano, 1995; Harvey, Leybourne & Newbold, 1997). Under scheme (a), predictors are selected for each estimation sample (i.e. in each estimation and forecast iteration), while under scheme (b) predictors are chosen over all estimation samples and are therefore the same in all estimation and forecast iterations.

Keeping the set of optimal predictors constant for each estimation sample as in scheme (b) results in improved forecasting performance vis-à-vis selecting the best predictors for each estimation sample as in scheme (a). Preselection, especially when it results in constant subsets of predictors over estimation samples (scheme (b)), leads to higher forecast accuracy, compared to the case in which no preselection of indicators is applied, i.e. the full dataset is employed for factor estimation and forecasting. Forecast gains compared to the benchmark reach 30% and 50% under scheme (a) and scheme (b), respectively. Under scheme (a), statistically significant gains are mainly concentrated in the middle of the horizon, while for one-quarter ahead forecasts the benefits of preselection are mostly associated with ADL combination forecasts. Under scheme (b), combinations of FAR or ADL models for GDP with predictors preselected either through hard or soft thresholding methods outperform both the benchmark and combinations based on the full panel of predictors, for all horizons considered.

Bottom-up GDP growth forecasts based on component forecasts computed using predictors specifically selected for each component under scheme (b), attain statistically significant gains for horizons of one and four quarters. Furthermore, for short horizons, bottom-up growth forecasts obtained from component FAR models with preselected predictors outperform bottom-up forecasts based on either all predictors or preselected predictors employed in component ADL models. In sum, the forecasting performance depends on (i) the horizon, with the largest gains achieved for $h = 4$, and, most importantly, (ii) whether the subsets of chosen predictors remain constant or change over the estimation samples. The technique employed for preselecting the predictors and the forecast combination applied (mean vs. DMSFE) are not found to considerably affect the forecasting performance.

Examining the forecasting performance at the GDP component level, we find that not all components are associated with significant gains over the simple AR benchmark (Table A4). Predictors do not seem to contain useful information for forecasting activity developments in the sectors of agriculture, information and communication, and real estate activities. For the remaining components, employing economic and financial indicators tends to improve the forecasting performance over the benchmark. More specifically, for the majority of components, gains are achieved with predictor preselection under scheme (b), i.e. when the sets of preselected predictors remain unchanged over estimation and forecasting iterations, a result also found for GDP. In the sectors of construction, as well as trade, transportation, accommodation and food services, GVA growth forecasts obtained from sets of preselected indicators outperform forecasts based on the full panel of predictors; preselection reduces the RMSFEs for all horizons, particularly in the case of the LASSO and LARS selection rules. Preselection under scheme (b) results in statistically significant gains over the benchmark for horizons up to four quarters in the industry sector; the gains, however, are close to those registered without preselection. For the rest of the services components (i.e. financial, professional, public, education, health and other services), some gains from preselection vis-à-vis the benchmark are found under scheme (b), but are not uniformly significant across selection methods, forecasting models and horizons; nevertheless, models with preselected predictors tend to result in lower RMSFE than those using the full panel. Finally, using preselected predictors for forecasting the net taxes component significantly improves on the benchmark and lowers the RMSFE compared to employing the full panel of predictors for forecasting. The highest gains from preselection over the benchmark (up to 51%) are attained for the construction, trade and net taxes components, while smaller gains (up to 17%) are found for the industry sector, as well as for financial and professional services, public administration, education, health and, other services. Forecasting with constant sets of preselected predictors improves the forecast accuracy for the majority of GDP components; the improvement is then reflected in the precision of the bottom-up GDP growth forecasts, especially for horizons up to four quarters.

4.2 Forecasting performance: crisis vs. normal times

This section examines whether the forecasting performance in the case of GDP growth varies depending on the state of the economy. As discussed in the previous section, using sets of preselected indicators in FAR and ADL models outperforms the AR benchmark significantly and results in higher forecasting gains vis-à-vis employing the full dataset of predictors. In this section, we investigate whether the RMSFEs obtained from the different predictor selection techniques and forecasting methods change substantially between crisis and normal times, and whether higher accuracy compared to the benchmark is maintained over the business cycle. Since the forecasting evaluation period is small, the analysis is descriptive in nature.

The performance of the methods considered for forecasting GDP growth is evaluated over two separate periods: (i) a crisis period, which includes the global financial crisis, the Greek debt crisis, and the ensuing crisis in Cyprus that peaked in 2013, and (ii) the rest of the sample, namely a period of steady growth (no crisis).[12, 13]

Figure 1 shows the RMSFEs for forecast combinations of FAR and ADL model forecasts relative to the AR benchmark. The figures report the relative RMSFEs over the full sample and the two sub-samples covering the crisis and no crisis periods; the RMSFEs are obtained using all predictors in the dataset for forecasting, as well as sets of indicators preselected through hard (F-test) and soft thresholding (LASSO and LARS) rules. We report the results for DMSFE combinations as this method penalises forecasts through past errors. Based on the findings concerning the forecasting performance over the full period (section 4.1), we focus on scheme (b), in which the sets of selected predictors remain constant over the estimation samples; scheme (b) results in considerably higher overall accuracy compared to scheme (a), namely the case of preselecting predictors in each estimation sample. The lower part of Figure 1 depicts the RMSFEs for bottom-up GDP growth forecasts obtained from aggregating component growth forecasts.

Looking at the crisis subsample, combinations of FAR or ADL model forecasts from preselected predictors outperform both the AR benchmark and the corresponding combinations that rely on the full panel of predictors; this result also holds for the full forecasting period. The largest gains from preselection during crisis reach 50% and 60% vis-à-vis the benchmark for horizons of one and four quarters, respectively, but decline to just above 10% for the eight-quarter horizon. The use of preselected predictors in FAR models to directly forecast GDP growth appears to be an optimal strategy for short-term forecasting ($h \leq 4$) during the crisis period; bottom-up growth forecasts based on FAR models and preselected predictors for GDP components also result in reliable short-term forecasts in the crisis period. The forecast accuracy for the longest horizon can be benefitted from predictor preselection during turbulent periods, but to a limited extent. During normal times (no crisis), predictor preselection leads to larger improvements in forecast precision for longer horizons, especially for $h = 8$, than for the one-quarter horizon. Applying indicator selection with respect to

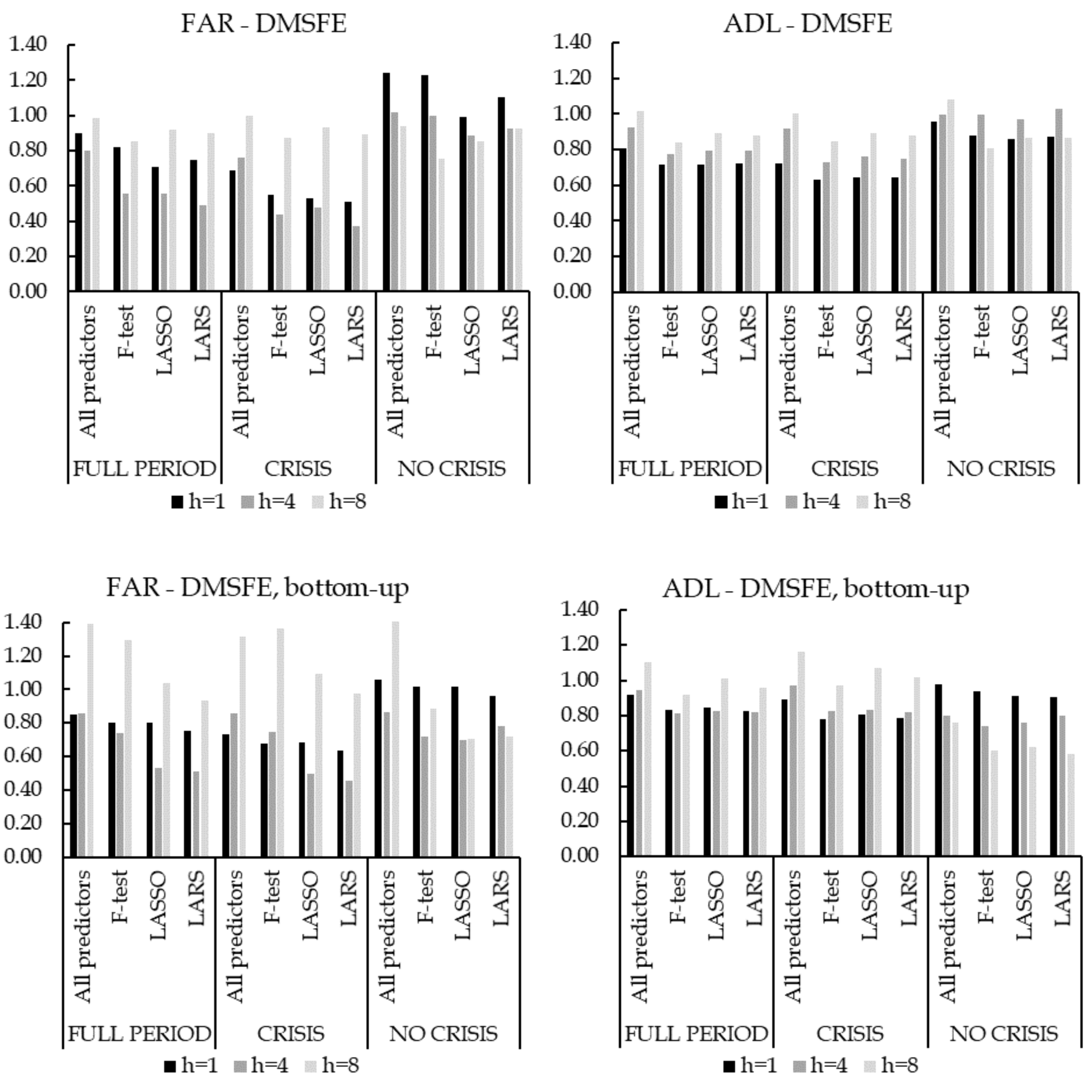
[12] The Greek debt crisis affected Cyprus mainly because of links between the banking systems in the two countries at the time.

[13] The crisis period includes the following quarters 2008Q3 – 2009Q4 and 2011Q3 – 2015Q1; these are quarters in which one of the following holds: (i) the quarter-on-quarter growth rate is negative, following a quarter with a negative quarter-on-quarter growth rate; (ii) the quarter-on-quarter growth rate is negative and the change in the unemployment rate is positive, following a quarter with a positive quarter-on-quarter growth rate, (iii) the quarter-on-quarter growth rate is positive and the change in the unemployment rate is positive, following a quarter with a negative quarter-on-quarter growth rate.

production-side components and computing bottom-up forecasts for GDP growth, results in the best performing forecasting strategy for the four- and eight-quarter horizons in normal times. Moreover, preselection improves forecasting precision for the shortest horizon in normal times, when the indicators are used in ADL models for GDP growth. Gains from preselection, vis-à-vis employing the full panel of predictors for forecasting, are larger in the crisis subsample than during normal times, averaging about 18% across methods and horizons; still, preselection leads to RMSFE reduction of about 14% on average in the no crisis subsample. Overall, preselection during the crisis period generates large forecast gains for horizons up to four quarters, while preselection in normal times benefits forecast accuracy for longer horizons.

FIGURE 1

Relative RMSFEs for different forecasting periods, GDP



5. Concluding remarks

The use of large datasets for macroeconomic forecasting has spurred research on the effects of preselecting predictors on the forecasting performance. A number of empirical studies document gains when a subset of predictors, as opposed to the full dataset, is employed for forecasting. This paper uses a dataset of 200 indicators and applies a hard thresholding method, based on the F-test, as well as the LASSO and LARS algorithms in order to select indicators from the dataset. The selected indicators are subsequently used for forecasting economic activity in Cyprus at the aggregate and sectoral levels.

The subsets of predictors are selected separately for each variable of interest, namely GDP and the 11 production-side components; also, the sets of selected predictors are allowed to vary over the forecast horizon. Predictor selection is carried out through two alternative schemes: (a) in each quarter, i.e. for each estimation sample, and (b) across all quarters, i.e. over all estimation samples. For the one-quarter horizon, the sets of selected predictors tend to contain a higher proportion of domestic as opposed to foreign predictors, while the opposite occurs for longer horizons. In the case of GDP, indicators from the group of business and consumer surveys are picked in high proportions for all horizons and by all thresholding techniques. For most of the production-side components, the sets of selected predictors contain mainly financial indicators, and, to a smaller degree, survey and real activity variables.

The performance of GDP and component growth forecasts computed from the selected sets of predictors is juxtaposed to that of forecasts constructed using the full dataset. Preselection leads to higher forecast accuracy compared to the case in which the full dataset is employed for forecasting, particularly when the sets of selected predictors remain constant over time. The thresholding technique employed for preselecting predictors is not found to substantially affect the forecasting performance. The benefits from preselecting predictors prior to forecasting GDP growth are enhanced during a crisis period, particularly for horizons up to four quarters. During normal times, forecast gains from preselection can also be attained vis-à-vis employing the full panel of predictors, especially when the forecasts are computed for longer horizons.

The results of the paper highlight the importance of timely available predictors, particularly survey and financial indicators, for forecasting economic activity in Cyprus. Narrowing down a large database to a smaller set of timely published indicators delivers a computationally fast tool for constructing nowcasts and monthly updates of forecasts. This tool can be particularly useful during times of turbulence in the economy.

Acknowledgements

The Economics Research Centre of the University of Cyprus acknowledges funding from the European Union, the Ministry of Finance and the University of Cyprus for conducting the project “Business and Consumer Surveys” in Cyprus.

Appendix

TABLE A1

List of dependent variables

Variable description	NACE Revision 2 classification	Abbreviation
Gross Domestic Product		GDP
Gross Value Added		GVA
1. Agriculture, forestry and fishing	A	AGR
2. Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply; water supply; sewerage, waste management and remediation activities	B, C, D, E	IND
3. Construction	F	CON
4. Wholesale and retail trade; repair of motor vehicles and motorcycles; transportation and storage; accommodation and food service activities	G, H, I	TRA
5. Information and communication	J	INF
6. Financial and insurance activities	K	FIN
7. Real estate activities	L	REA
8. Professional, scientific and technical activities; administrative and support service activities	M, N	PRO
9. Public administration and defence; compulsory social security; education; human health and social work activities	O, P, Q	PEH
10. Arts, entertainment and recreation, other services	R, S, T, U	OTH
Taxes less subsidies on products		TAX

Note: The variables refer to seasonally adjusted, chain-linked volume measures; the data were obtained from the Statistical Service of Cyprus.

TABLE A2

List of predictors in the dataset

Variable description	Details		
	A	B	C
Registration of motor vehicles (number of passenger cars)	D	M	3
Registration of motor vehicles (number of all vehicles)	D	M	3
Volume index of retail trade - NACE 47, (2015=100)	D	M	3
Volume index of retail trade - NACE 47 excl. 47.3, (2015=100)	D	M	3
Building permits authorised (actual number)	D	M	3
Building permits authorised (value), deflated	D	M	3
Volume index of manufacturing production, (2015=100)	D	M	3
Index of industrial production, (2015=100)	D	M	3
Total imports/arrivals (c.i.f.), deflated	D	M	3
Total exports/dispatches (f.o.b.), deflated	D	M	3
Tourist arrivals in Cyprus	D	M	3
Tourist arrivals in Cyprus from the United Kingdom (usual residency)	D	M	3
Tourist arrivals in Cyprus from Germany (usual residency)	D	M	3
Tourist arrivals in Cyprus from Russia (usual residency)	D	M	3
Tourist arrivals in Cyprus from Greece (usual residency)	D	M	3
Tourism revenues in Cyprus (act €mn), deflated	D	M	3
Local cards used in Cyprus (sales value), deflated	D	M	3
Foreign cards used in Cyprus (sales value), deflated	D	M	3
Total number of property sale contracts (registered contracts, Department of Lands and Surveys)	D	M	3
Total number of property transfers (number of parcels, Department of Lands and Surveys)	D	M	3
Total value of property transfers (accepted amount, Department of Lands and Surveys), deflated	D	M	3
Registration of new companies in Cyprus (number)	D	M	3
VAT on products, deflated	D	M	3
Total registered unemployed (actual number)	D	M	3
Registered unemployed - Agriculture, forestry and fishing (actual number)	D	M	3
Registered unemployed - Manufacturing (actual number)	D	M	3
Registered unemployed - Construction (actual number)	D	M	3
Registered unemployed - Wholesale & retail trade (actual number)	D	M	3
Registered unemployed - Restaurants & hotels (actual number)	D	M	3
Registered unemployed - Transport, storage & communication (actual number)	D	M	3
Registered unemployed - Finance, insurance, real estates & business services (actual number)	D	M	3
Registered unemployed - Newcomers (actual number)	D	M	3
Unemployment rate	D	M	2

TABLE A2 (continued)

Variable description	Details		
	A	B	C
Cyprus Stock Exchange, All Share General Composite Index	D	M	3
Cyprus Stock Exchange, All Share General Composite Index, historical volatility	D	M	1
Cyprus Stock Exchange, FTSE/CySE 20 Index	D	M	3
Cyprus Stock Exchange, Hotels Index	D	M	3
Cyprus Stock Exchange, Investment Companies Index	D	M	3
Interest rates on deposits of non-financial corporations (new business), overnight	D	M	2
Interest rates on deposits of non-financial corporations (new business), maturity up to 1 year	D	M	2
Interest rates on deposits of households (new business), overnight	D	M	2
Interest rates on deposits of households (new business), redeemable at notice up to 3 months	D	M	2
Interest rates on deposits of households (outstanding amounts), redeemable at notice over 3 months	D	M	2
Interest rates on deposits of households (new business), maturity up to 1 year	D	M	2
Interest rates on loans of non-financial corporations (new business), bank overdrafts	D	M	2
Interest rates on loans of non-financial corporations (new business) - up to EUR 1 mn - floating rate up to 1 year initial fixation	D	M	2
Interest rates on loans of non-financial corporations (new business) - over EUR 1 mn - floating rate up to 1 year initial fixation	D	M	2
Interest rates on loans of households (new business), consumer credit, floating rate up to 1 year initial fixation	D	M	2
Interest rates on loans of households (new business), lending for house purchase, floating rate up to 1 year initial fixation	D	M	2
Interest rates on loans of households (new business), other lending, floating rate up to 1 year initial fixation	D	M	2
Interest rates on loans of households (outstanding amounts), maturity up to 1 year	D	M	2
Loans to non-MFIs (outstanding amounts), total, annual growth rates	D	M	1
Loans to non-MFIs (outstanding amounts), domestic residents, annual growth rates	D	M	1
Deposits of non-MFIs held with MFIs (outstanding amounts), total, annual growth rates	D	M	1
Deposits of non-MFIs held with MFIs (outstanding amounts), domestic residents, annual growth rates	D	M	1
Loan-to-deposit ratio: total loans to total deposits	D	M	2
Volumes of new loans to euro area non-financial corporations & euro area households, pure new loans (€mn)	D	M	3
Volumes of new loans to euro area non-financial corporations, pure new loans (€mn)	D	M	3
Volumes of new loans to euro area households, pure new loans (€mn)	D	M	3
Volumes of new loans to euro area households, pure new loans for house purchase (€mn)	D	M	3
Harmonised Index of Consumer Prices (HICP), (2015=100)	D	M	3
Consumer Price Index, (2015=100)	D	M	3
Price Index of construction materials, (2015=100)	D	M	3
Economic Sentiment Indicator	D	M	2
Economic Sentiment Indicator (Economics Research Centre)	D	M	2
Employment Expectations Indicator	D	M	2
Employment Expectations Indicator (Economics Research Centre)	D	M	2
Services Confidence Indicator	D	M	2
Services, Assessments of business situation over the past 3 months, balance	D	M	2
Services, Evolution of demand over the past 3 months, balance	D	M	2
Services, Demand expectations over the next 3 months, balance	D	M	2
Services, Evolution of employment over the past 3 months, balance	D	M	2
Services, Employment expectations over the next 3 months, balance	D	M	2
Services, Selling price expectations over the next 3 months	D	M	2
Services, Hotels & Restaurants Confidence Indicator (NACE 55, 56)	D	M	2
Services, Financial Services Confidence Indicator (NACE 64, 65, 66)	D	M	2
Services, Capacity utilisation index	D	Q	1
Retail Trade Confidence Indicator	D	M	2
Retail Trade, Assessments of business activity (sales) over the past 3 months, balance	D	M	2
Retail Trade, Assessments of volume of stock currently hold, balance	D	M	2
Retail Trade, Intentions of placing orders over the next 3 months, balance	D	M	2
Retail Trade, Sales expectations over the next 3 months, balance	D	M	2
Retail Trade, Employment expectations over the next 3 months, balance	D	M	2
Retail Trade, Selling price expectations over the next 3 months, balance	D	M	2
Construction Confidence Indicator	D	M	2
Construction, Assessments of building activity over the past 3 months, balance	D	M	2
Construction, Assessments of current overall order books, balance	D	M	2
Construction, Employment expectations over the next 3 months, balance	D	M	2
Construction, Selling price expectations over the next 3 months, balance	D	M	2
Construction, Operating time ensured by current backlog, months	D	Q	1
Industry Confidence Indicator	D	M	2
Industry, Assessment of production over the past 3 months, balance	D	M	2
Industry, Assessment of current order books, balance	D	M	2
Industry, Assessment of current export order books, balance	D	M	2
Industry, Assessment of current stock of finished products, balance	D	M	2
Industry, Production expectations over the next 3 months, balance	D	M	2
Industry, Selling price expectations for the next 3 months, balance	D	M	2
Industry, Employment expectations for the next 3 months, balance	D	M	2
Industry, Capacity utilisation index	D	Q	1
Consumer Confidence Indicator	D	M	2
Consumers, Financial situation of household over the last 12 months, balance	D	M	2
Consumers, Financial situation of household over the next 12 months, balance	D	M	2
Consumers, General economic situation over the last 12 months, balance	D	M	2
Consumers, General economic situation over the next 12 months, balance	D	M	2

TABLE A2 (continued)

Variable description	Details		
	A	B	C
Consumers, Price trends over the last 12 months, balance	D	M	2
Consumers, Price trends over the next 12 months, balance	D	M	2
Consumers, Unemployment expectations over the next 12 months, balance	D	M	2
Consumers, Major purchases at present, balance	D	M	2
Consumers, Major purchases over the next 12 months, balance	D	M	2
Consumers, Savings at present, balance	D	M	2
Consumers, Savings over the next 12 months, balance	D	M	2
Consumers, Statement on financial situation of household, balance	D	M	2
Consumers, Intention to buy a car within the next 12 months, balance	D	Q	2
Consumers, Purchase or build a home within the next 12 months, balance	D	Q	2
Consumers, Home improvements over the next 12 months, balance	D	Q	2
Uncertainty indicator, ex ante employment and price expectations (business surveys)	D	M	1
Uncertainty indicator, ex post, expectation errors (business surveys)	D	M	1
EU27, Volume index of manufacturing production, (2015=100)	F	M	3
EA19, Volume index of manufacturing production, (2015=100)	F	M	3
EU27, Retail - except of motor vehicle and motorcycles, Index of deflated turnover, (2015=100)	F	M	3
EA19, Retail - except of motor vehicle and motorcycles, Index of deflated turnover, (2015=100)	F	M	3
EA, Real GDP growth expectations, current calendar year (ECB Survey of Professional Forecasters)	F	Q	1
EA, Real GDP growth expectations, next calendar year (ECB Survey of Professional Forecasters)	F	Q	1
EU, Unemployment rate	F	M	2
EA, Unemployment rate	F	M	2
UK, Unemployment rate	F	M	2
Russia, Unemployment rate	F	M	2
EA, Unemployment rate expectations, current calendar year (ECB Survey of Professional Forecasters)	F	Q	1
EA, Unemployment rate expectations, next calendar year (ECB Survey of Professional Forecasters)	F	Q	1
Europe, 3-month EURIBOR	F	M	2
Europe, 6-month EURIBOR	F	M	2
Europe, 12-month EURIBOR	F	M	2
European Central Bank deposit rate	F	M	2
European Central Bank lending rate	F	M	2
Germany, 10-year Government Benchmark Bond Yield (DE10)	F	M	2
Germany, 3-month Treasury Bill Yield	F	M	2
France, 10-year Government Bond Yield (FR10)	F	M	2
France, 3-month Treasury Bill Yield	F	M	2
Italy, 10-year Government Benchmark Bond Yield (IT10)	F	M	2
Italy, 3-month Treasury Bill Yield	F	M	2
Spain, 10-year Government Benchmark Bond Yield (ES10)	F	M	2
Spain, 3-month Treasury Bill Yield	F	M	2
Greece, 10-year Government Note Yield (EL10)	F	M	2
Greece, 3-month Treasury Bill Yield	F	M	2
UK, 10-year Government Bond Yield (UK10)	F	M	2
UK, 3-month Treasury Bill Yield	F	M	2
Spread FR10 and DE10	F	M	1
Spread IT10 and DE10	F	M	1
Spread ES10 and DE10	F	M	1
Spread EL10 and DE10	F	M	1
Euro Dow Jones Euro Stoxx 50 Price Index, Euro area (changing composition)	F	M	3
Euro Dow Jones Euro Stoxx Price Index, Euro area (changing composition)	F	M	3
FTSE 100 Price Index	F	M	3
DAX 30 Performance Price Index	F	M	3
CAC 40 Price Index	F	M	3
ATHEX Composite Price Index	F	M	3
S&P 500 Composite Price Index	F	M	3
S&P 100 Price Index	F	M	3
NYSE Composite Price Index	F	M	3
MICEX Share Price Index	F	M	3
Nikkei 225 Stock Average Index	F	M	3
CBOE S&P 500 Volatility Index	F	M	1
EURO STOXX 50 Price Index, historical volatility (based on historical data of EURO STOXX 50 Price Index)	F	M	1
S&P500 Index, historical volatility (based on historical data of S&P500 Index)	F	M	1
Exchange rate, euro against the British pound	F	M	3
Exchange rate, euro against the Russian ruble	F	M	3
Exchange rate, euro against the US dollar	F	M	3
Exchange rate, euro against the Chinese renminbi yuan	F	M	3
EA, Harmonised Index of Consumer Prices (HICP), (2015=100)	F	M	3
EU, Harmonised Index of Consumer Prices (HICP), (2015=100)	F	M	3
UK, Harmonised Index of Consumer Prices (HICP), (2015=100)	F	M	3
Russia, Consumer Price Index	F	M	3
Brent Crude Oil (€)	F	M	3
West Texas Intermediate Oil Price (€/Barrel)	F	M	3
Crude Oil Futures (€) - Futures Contracts	F	M	3

TABLE A2 (continued)

Variable description	Details		
	A	B	C
Gold Bullion Price - New York (€/Ounce)	F	M	3
Silver Cash Price (€/Ounce)	F	M	3
Wheat Cash Price (€/Bushel)	F	M	3
Euro area 17 (fixed composition) ECB Commodity Price index, Euro denominated, use-weighted, Food	F	M	3
Euro area 17 (fixed composition) ECB Commodity Price index, Euro denominated, use-weighted, Non-food	F	M	3
Euro area 17 (fixed composition) ECB Commodity Price index, Euro denominated, use-weighted, Total non-energy commodity	F	M	3
EA, HICP inflation expectations, current calendar year (ECB Survey of Professional Forecasters)	F	Q	1
EA, HICP inflation expectations, next calendar year (ECB Survey of Professional Forecasters)	F	Q	1
Global Economic Policy Uncertainty Index	F	M	1
EU, Economic Sentiment Indicator	F	M	2
EA, Economic Sentiment Indicator	F	M	2
UK, Economic Sentiment Indicator	F	M	2
Greece, Economic Sentiment Indicator	F	M	2
EU, Employment Expectations Indicator	F	M	2
EA, Employment Expectations Indicator	F	M	2
EU, Services Confidence Indicator	F	M	2
EU, Industry Confidence Indicator	F	M	2
EU, Retail Trade Confidence Indicator	F	M	2
EU, Construction Confidence Indicator	F	M	2
EU, Consumer Confidence Indicator	F	M	2
EA, Services Confidence Indicator	F	M	2
EA, Retail Trade Confidence Indicator	F	M	2
EA, Construction Confidence Indicator	F	M	2
EA, Industry Expectations Indicator	F	M	2
EA, Consumer Confidence Indicator	F	M	2

Notes: The column “Details” is read as follows: in column A the symbols “D” and “F” denote domestic and foreign/international indicators, respectively; in column B the symbols “M” and “Q” denote monthly and quarterly data frequency, respectively; in column C the transformation codes 1, 2 and 3 represent a series in levels, first difference of levels, and first difference of the logarithm of levels, respectively.

TABLE A3

Percentages of selected predictors by type, and cut-off frequencies, production-side components

Horizon (quarters)	1			4			8		
Predictor selection method	F-test	LASSO	LARS	F-test	LASSO	LARS	F-test	LASSO	LARS
AGR									
Types of predictors									
Domestic	71	65	52	36	29	25	40	35	25
Foreign/international	29	35	48	64	71	75	60	65	76
By category									
Real activity and labour market	5	25	19	36	29	30	20	25	40
Financial	14	30	33	14	38	30	50	50	35
Prices	19	10	19	32	14	20	20	10	10
Business and consumer surveys	62	35	29	18	19	20	10	15	15
Cut-off frequency (%)	38	29	28	26	20	23	35	23	18
IND									
Types of predictors									
Domestic	62	76	65	45	48	50	55	43	45
Foreign/international	38	24	35	55	52	50	45	57	55
By category									
Real activity and labour market	38	33	25	5	10	0	25	14	5
Financial	10	38	35	45	52	45	40	57	60
Prices	38	5	15	30	19	30	5	10	10
Business and consumer surveys	14	24	25	20	19	25	30	19	25
Cut-off frequency (%)	33	29	31	38	29	29	34	28	29
CON									
Types of predictors									
Domestic	45	70	76	50	70	62	45	45	48
Foreign/international	55	30	24	50	30	38	55	55	52
By category									
Real activity and labour market	45	30	43	35	55	38	20	35	29
Financial	25	30	24	40	35	48	50	50	52
Prices	5	5	5	10	5	5	15	10	10
Business and consumer surveys	25	35	29	15	5	10	15	5	10
Cut-off frequency (%)	39	28	24	36	29	28	40	28	23
TRA									
Types of predictors									
Domestic	40	65	38	15	30	30	35	43	40
Foreign/international	60	35	62	85	70	70	65	57	60
By category									
Real activity and labour market	10	40	29	0	10	10	10	17	20
Financial	40	25	33	55	40	40	55	35	40
Prices	0	5	10	5	30	25	15	22	20
Business and consumer surveys	50	30	29	40	20	25	20	26	20
Cut-off frequency (%)	40	33	36	46	25	23	45	32	34
INF									
Types of predictors									
Domestic	67	71	57	38	55	41	35	48	35
Foreign/international	33	29	43	62	45	59	65	52	65
By category									
Real activity and labour market	10	19	5	10	5	9	20	13	30
Financial	19	29	52	43	35	50	40	39	40
Prices	14	14	10	10	15	14	20	17	20
Business and consumer surveys	57	38	33	38	45	27	20	30	10
Cut-off frequency (%)	35	26	31	41	28	28	45	25	26
Types of predictors									
FIN									
Types of predictors									
Domestic	40	61	60	30	48	35	20	48	33
Foreign/international	60	39	40	70	52	65	80	52	67
By category									
Real activity and labour market	45	43	50	10	29	20	10	30	24
Financial	35	17	20	60	24	35	80	39	38
Prices	15	22	15	0	10	10	0	13	10
Business and consumer surveys	5	17	15	30	38	35	10	17	29
Cut-off frequency (%)	40	25	24	39	28	25	48	28	29

TABLE A3 (continued)

Horizon (quarters)	1			4			8		
Predictor selection method	F-test	LASSO	LARS	F-test	LASSO	LARS	F-test	LASSO	LARS
REA									
Types of predictors									
Domestic	30	52	45	18	36	30	48	60	52
Foreign/international	70	48	55	82	64	70	52	40	48
By category									
Real activity and labour market	10	10	10	18	14	15	29	20	22
Financial	65	38	45	50	41	40	48	40	39
Prices	15	24	20	18	23	30	5	15	9
Business and consumer surveys	10	29	25	14	23	15	19	25	30
Cut-off frequency (%)	33	24	29	35	26	26	29	22	17
PRO									
Types of predictors									
Domestic	35	50	57	20	20	25	19	30	29
Foreign/international	65	50	43	80	80	75	81	70	71
By category									
Real activity and labour market	30	20	29	20	15	20	14	20	24
Financial	30	35	33	50	50	50	76	45	43
Prices	0	0	0	5	10	10	0	5	10
Business and consumer surveys	40	45	38	25	25	20	10	30	24
Cut-off frequency (%)	40	31	24	36	32	25	43	26	18
PEH									
Types of predictors									
Domestic	48	52	50	38	45	35	60	55	50
Foreign/international	52	48	50	62	55	65	40	45	50
By category									
Real activity and labour market	13	10	10	33	5	25	50	20	20
Financial	17	19	35	14	45	35	40	40	35
Prices	0	10	5	5	0	10	0	5	10
Business and consumer surveys	70	62	50	48	50	30	10	35	35
Cut-off frequency (%)	31	28	28	32	29	26	38	28	25
OTH									
Types of predictors									
Domestic	40	55	45	25	40	30	33	38	25
Foreign/international	60	45	55	75	60	70	66	62	75
By category									
Real activity and labour market	15	15	15	15	20	15	33	19	25
Financial	25	30	25	40	50	45	43	48	50
Prices	35	30	40	30	15	20	14	19	10
Business and consumer surveys	25	25	20	15	15	20	10	14	15
Cut-off frequency (%)	42	32	33	38	28	30	42	22	26
TAX									
Types of predictors									
Domestic	50	75	80	29	35	29	35	40	30
Foreign/international	50	25	20	71	65	71	65	60	70
By category									
Real activity and labour market	35	30	50	5	25	24	5	5	10
Financial	20	20	20	38	20	29	45	40	45
Prices	0	5	0	19	20	19	20	15	10
Business and consumer surveys	45	45	30	38	35	29	30	40	35
Cut-off frequency (%)	36	31	29	38	33	30	45	29	32

Note: The cut-off frequency shows the lowest frequency with which an indicator should be chosen across sample iterations by a given method in order to be included in the set of selected predictors.

TABLE A4

Relative RMSFE (benchmark AR), production-side components

Horizon (quarters)	1				4				8			
Predictor selection method	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors
AGR												
Benchmark AR(1): RMSFE	2.49				8.55				9.13			
Scheme (a): each *t*												
FAR - mean	1.21	1.28	1.23	1.09	1.29	1.28	1.27	1.15	1.20	1.44	1.42	1.14
FAR - DMSFE	1.21	1.29	1.23	1.09	1.30	1.27	1.27	1.14	1.20	1.44	1.42	1.14
ADL - mean	1.05	1.05	1.05	1.02	1.17	1.08	1.09	1.07	1.10	1.16	1.18	1.05
ADL - DMSFE	1.08	1.05	1.03	1.02	1.12	1.11	1.13	1.08	1.18	1.13	1.20	1.05
Scheme (b): all *t*												
FAR - mean	1.05	1.12	1.10	1.09	1.14	0.96	1.09	1.15	1.24	1.11	1.02	1.14
FAR - DMSFE	1.06	1.10	1.10	1.09	1.12	0.96	1.08	1.14	1.22	1.13	1.02	1.14
ADL - mean	0.99	1.00	1.01	1.02	1.04	1.04	1.03	1.07	1.01	1.03	1.06	1.05
ADL - DMSFE	0.99	1.01	1.01	1.02	1.03	1.04	1.03	1.08	1.04	1.06	1.08	1.05
IND												
Benchmark AR(1): RMSFE	1.87				6.13				8.21			
Scheme (a): each *t*												
FAR - mean	1.00	1.03	0.97	**0.88**	1.04	0.94	0.94	**0.88**	1.13	1.06	1.14	1.12
FAR - DMSFE	1.00	1.01	0.97	**0.88**	1.04	0.94	0.94	**0.87**	1.13	1.06	1.14	1.13
ADL - mean	**0.91**	**0.94**	**0.93**	**0.95**	0.93	**0.94**	**0.93**	0.99	1.01	0.99	0.95	1.04
ADL - DMSFE	**0.91**	**0.94**	**0.93**	**0.95**	1.02	1.02	1.01	0.99	0.95	0.95	0.88	1.03
Scheme (b): all *t*												
FAR - mean	**0.87**	**0.93**	**0.84**	**0.88**	0.96	0.86	0.94	**0.88**	0.99	1.00	1.06	1.12
FAR - DMSFE	**0.87**	**0.93**	**0.83**	**0.88**	0.97	0.86	0.94	**0.87**	1.00	1.00	1.07	1.13
ADL - mean	**0.90**	**0.93**	**0.93**	**0.95**	**0.88**	**0.91**	**0.93**	0.99	0.93	0.98	0.98	1.04
ADL - DMSFE	**0.89**	**0.91**	**0.91**	**0.95**	**0.87**	**0.90**	**0.92**	0.99	0.86	0.91	0.90	1.03
CON												
Benchmark AR(1): RMSFE	5.09				15.57				18.48			
Scheme (a): each *t*												
FAR - mean	1.21	1.15	1.16	0.91	1.02	1.06	1.03	1.04	0.95	0.94	1.04	1.13
FAR - DMSFE	1.20	1.17	1.13	0.90	1.02	1.06	1.03	1.05	0.95	0.94	1.04	1.13
ADL - mean	**0.91**	**0.93**	**0.93**	**0.91**	0.94	0.89	0.91	**0.88**	1.02	0.99	0.99	1.00
ADL - DMSFE	0.94	1.00	0.99	0.93	0.95	0.97	0.91	**0.88**	1.03	0.98	1.03	1.00
Scheme (b): all *t*												
FAR - mean	0.88	**0.77**	**0.85**	0.91	1.09	**0.74**	**0.67**	1.04	1.06	**0.73**	**0.68**	1.13
FAR - DMSFE	**0.86**	**0.78**	**0.85**	0.90	1.09	**0.74**	**0.67**	1.05	1.08	**0.73**	**0.68**	1.13
ADL - mean	**0.86**	**0.89**	**0.88**	**0.91**	**0.84**	**0.83**	**0.86**	**0.88**	0.96	0.96	0.96	1.00
ADL - DMSFE	**0.89**	**0.90**	**0.89**	0.93	**0.84**	**0.84**	**0.85**	**0.88**	0.93	0.94	0.94	1.00
TRA												
Benchmark AR(1): RMSFE	1.60				5.21				6.04			
Scheme (a): each *t*												
FAR - mean	0.99	0.90	0.92	0.99	0.87	0.77	0.71	1.01	1.00	1.07	1.10	1.15
FAR - DMSFE	0.98	0.90	0.91	0.99	0.87	0.77	0.71	1.01	1.00	1.07	1.11	1.15
ADL - mean	**0.89**	**0.88**	**0.86**	**0.93**	0.86	**0.81**	**0.79**	**0.93**	1.01	0.97	0.98	**0.98**
ADL - DMSFE	**0.87**	**0.88**	**0.85**	**0.93**	0.90	**0.83**	**0.81**	**0.92**	1.16	0.98	1.01	0.98
Scheme (b): all *t*												
FAR - mean	**0.90**	**0.79**	**0.80**	0.99	0.89	**0.68**	0.67	1.01	0.89	0.93	0.86	1.15
FAR - DMSFE	**0.91**	**0.78**	**0.80**	0.99	0.89	**0.69**	0.67	1.01	0.89	0.93	0.86	1.15
ADL - mean	**0.86**	**0.85**	**0.82**	**0.93**	**0.85**	**0.80**	**0.75**	**0.93**	0.85	**0.88**	**0.86**	**0.98**
ADL - DMSFE	**0.86**	**0.84**	**0.81**	**0.93**	**0.85**	**0.78**	**0.74**	**0.92**	0.84	**0.88**	**0.87**	0.98

TABLE A4 (continued)

Horizon (quarters)	1				4				8			
Predictor selection method	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors
INF												
Benchmark AR(1): RMSFE		5.34				11.29				11.13		
Scheme (a): each *t*												
FAR - mean	1.13	1.14	1.14	1.05	1.17	1.40	1.23	1.24	1.13	1.36	1.39	1.00
FAR - DMSFE	1.14	1.14	1.12	1.05	1.16	1.40	1.23	1.18	1.13	1.36	1.39	1.00
ADL - mean	1.08	1.06	1.07	1.05	1.04	1.01	1.00	1.01	1.06	1.03	1.08	1.00
ADL - DMSFE	1.06	1.08	1.07	1.04	1.03	1.04	1.01	1.02	1.10	1.13	1.32	1.01
Scheme (b): all *t*												
FAR - mean	1.09	1.08	1.01	1.05	1.09	0.90	0.93	1.24	1.26	0.91	0.94	1.00
FAR - DMSFE	1.09	1.05	1.01	1.05	1.08	0.91	0.93	1.18	1.25	0.91	0.94	1.00
ADL - mean	1.04	1.04	1.04	1.05	0.93	0.96	0.93	1.01	0.98	0.98	0.98	1.00
ADL - DMSFE	1.04	1.04	1.03	1.04	0.94	0.98	0.93	1.02	0.98	0.99	1.00	1.01
FIN												
Benchmark AR(1): RMSFE		4.03				9.05				10.31		
Scheme (a): each *t*												
FAR - mean	1.80	1.29	2.19	1.08	1.26	1.18	1.11	1.21	1.22	1.05	1.11	1.28
FAR - DMSFE	1.72	1.31	2.10	1.13	1.26	1.18	1.11	1.21	1.23	1.05	1.11	1.27
ADL - mean	0.98	0.99	0.99	**0.97**	1.03	0.98	0.97	1.00	1.02	1.03	1.06	1.00
ADL - DMSFE	0.98	1.06	0.99	**0.94**	1.07	1.02	0.98	1.00	1.00	0.98	1.04	1.00
Scheme (b): all *t*												
FAR - mean	1.03	1.11	1.11	1.08	1.23	1.12	1.02	1.21	1.20	1.10	1.14	1.28
FAR - DMSFE	1.04	1.11	1.10	1.13	1.20	1.12	1.02	1.21	1.20	1.10	1.13	1.27
ADL - mean	**0.94**	**0.93**	**0.94**	**0.97**	0.99	**0.98**	**0.95**	1.00	0.97	1.00	0.99	1.00
ADL - DMSFE	0.93	**0.92**	**0.90**	**0.94**	0.96	**0.94**	**0.92**	1.00	0.96	0.98	0.96	1.00
REA												
Benchmark AR(1): RMSFE		2.46				3.96				4.60		
Scheme (a): each *t*												
FAR - mean	1.05	0.98	0.95	1.00	1.10	1.01	1.05	1.05	1.11	1.21	1.16	1.05
FAR - DMSFE	1.05	0.98	0.95	1.01	1.10	1.01	1.05	1.04	1.10	1.21	1.16	1.05
ADL - mean	1.02	1.02	1.02	1.04	1.02	1.01	1.01	1.06	1.06	1.04	1.03	1.03
ADL - DMSFE	1.02	1.03	1.02	1.04	1.01	1.02	1.03	1.05	1.04	1.02	1.02	1.03
Scheme (b): all *t*												
FAR - mean	1.01	0.98	0.99	1.00	1.01	0.79	0.84	1.05	1.00	1.01	0.92	1.05
FAR - DMSFE	1.01	0.98	0.99	1.01	1.00	0.80	0.84	1.04	1.00	1.01	0.92	1.05
ADL - mean	1.02	1.02	1.02	1.04	1.02	1.02	1.02	1.06	1.02	1.02	1.03	1.03
ADL - DMSFE	1.02	1.02	1.02	1.04	1.02	1.02	1.02	1.05	1.02	1.02	1.03	1.03
PRO												
Benchmark AR(1): RMSFE		1.43				4.48				5.15		
Scheme (a): each *t*												
FAR - mean	1.06	1.17	1.24	1.01	1.34	1.31	1.36	1.27	0.98	1.05	1.11	1.26
FAR - DMSFE	1.06	1.14	1.21	1.01	1.34	1.31	1.36	1.27	0.98	1.06	1.11	1.26
ADL - mean	0.98	0.95	0.96	0.97	1.04	0.97	0.99	0.98	0.99	0.97	0.97	0.99
ADL - DMSFE	1.00	0.96	0.94	0.96	1.07	1.09	1.09	0.97	1.03	0.98	0.98	1.00
Scheme (b): all *t*												
FAR - mean	1.07	1.03	1.04	1.01	1.12	1.08	1.08	1.27	1.34	0.86	0.93	1.26
FAR - DMSFE	1.07	1.03	1.04	1.01	1.11	1.08	1.09	1.27	1.33	0.86	0.93	1.26
ADL - mean	0.94	0.94	0.92	0.97	0.90	**0.91**	0.92	0.98	0.93	**0.92**	**0.93**	0.99
ADL - DMSFE	0.92	0.92	0.92	0.96	0.89	0.92	0.92	0.97	0.96	0.95	0.96	1.00

TABLE A4 (continued)

Horizon (quarters)	1				4				8			
Predictor selection method	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors	F-test	LASSO	LARS	All predictors
PEH												
Benchmark AR(1): RMSFE		0.71				2.27				2.93		
Scheme (a): each *t*												
FAR - mean	1.16	1.08	1.09	1.01	1.15	1.18	1.21	1.02	1.08	1.11	1.13	0.89
FAR - DMSFE	1.16	1.08	1.08	1.01	1.15	1.17	1.21	1.03	1.08	1.11	1.13	0.89
ADL - mean	0.99	0.99	0.98	1.00	1.05	1.01	1.04	1.03	0.97	1.03	1.00	1.06
ADL - DMSFE	1.02	1.00	0.97	1.00	1.11	0.96	1.04	1.03	0.93	**0.96**	0.99	1.05
Scheme (b): all *t*												
FAR - mean	1.00	0.98	1.05	1.01	**0.81**	**0.64**	1.13	1.02	0.92	0.92	1.02	0.89
FAR - DMSFE	1.01	0.98	1.05	1.01	**0.80**	**0.64**	1.15	1.03	0.92	0.92	1.02	0.89
ADL - mean	0.94	0.98	0.96	1.00	0.89	0.98	0.98	1.03	**0.88**	1.02	1.01	1.06
ADL - DMSFE	0.94	0.98	0.95	1.00	**0.88**	0.98	0.97	1.03	**0.87**	0.99	0.98	1.05
OTH												
Benchmark AR(1): RMSFE		1.91				7.32				8.55		
Scheme (a): each *t*												
FAR - mean	0.98	0.96	1.09	1.11	1.13	1.12	1.17	1.01	1.20	1.26	1.13	1.12
FAR - DMSFE	0.96	0.95	1.11	1.06	1.13	1.12	1.18	1.02	1.19	1.26	1.12	1.12
ADL - mean	1.02	0.99	1.00	1.02	1.02	1.04	1.04	1.05	1.03	1.03	1.00	0.99
ADL - DMSFE	1.03	1.00	1.01	1.01	1.00	1.05	1.05	1.04	1.03	1.12	1.08	0.99
Scheme (b): all *t*												
FAR - mean	0.98	1.05	1.01	1.11	1.16	0.89	0.85	1.01	1.16	1.20	1.25	1.12
FAR - DMSFE	0.96	1.04	0.97	1.06	1.16	0.90	0.85	1.02	1.16	1.20	1.25	1.12
ADL - mean	**0.97**	**0.98**	**0.97**	1.02	0.99	1.03	1.02	1.05	1.02	1.00	1.01	0.99
ADL - DMSFE	**0.95**	**0.97**	**0.96**	1.01	**0.98**	1.02	1.00	1.04	1.01	1.00	1.00	0.99
TAX												
Benchmark AR(1): RMSFE		1.33				4.02				4.82		
Scheme (a): each *t*												
FAR - mean	**0.85**	0.91	**0.76**	**0.83**	0.73	**0.70**	**0.66**	0.81	1.01	1.07	1.06	1.01
FAR - DMSFE	0.87	0.94	**0.77**	**0.81**	0.72	**0.69**	**0.66**	0.80	1.01	1.07	1.06	1.01
ADL - mean	**0.77**	**0.83**	**0.82**	**0.85**	0.87	**0.81**	**0.84**	**0.93**	1.00	0.98	0.97	1.00
ADL - DMSFE	**0.74**	**0.78**	**0.69**	**0.80**	0.77	**0.86**	0.89	**0.92**	1.00	1.08	**0.97**	1.00
Scheme (b): all *t*												
FAR - mean	**0.72**	**0.72**	0.94	**0.83**	**0.50**	**0.53**	**0.50**	0.81	0.96	**0.92**	**0.89**	1.01
FAR - DMSFE	**0.72**	**0.71**	0.88	**0.81**	**0.50**	**0.53**	**0.49**	0.80	0.95	**0.92**	**0.89**	1.01
ADL - mean	**0.73**	**0.79**	**0.78**	**0.85**	**0.75**	**0.80**	**0.80**	**0.93**	**0.87**	**0.89**	0.88	1.00
ADL - DMSFE	**0.68**	**0.70**	**0.70**	**0.80**	**0.75**	**0.78**	**0.78**	**0.92**	0.88	0.88	0.88	1.00

See notes to Table 3.

References

Baffigi, A., Golinelli, R. and Parigi, G., (2004), "Bridge models to forecast the euro area GDP", *International Journal of Forecasting*, 20(3): 447-460.

Bai, J., and Ng, S., (2002), "Determining the number of factors in approximate factor models", *Econometrica,* 70(1): 191-221.

Bai, J., and Ng, S., (2006), "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions", *Econometrica*, 74(4): 1133-1150.

Bai, J., and Ng, S., (2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics*, 146(2): 304-317.

Banbura, M. and Runstler, G., (2011), "A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP", *International Journal of Forecasting*, 27(2): 333-346.

Boivin, J., and Ng, S., (2006), "Are more data always better for factor analysis?", *Journal of Econometrics*, 132(1): 169-194.

Bulligan, G., Marcellino, M., and Venditti, F., (2015), "Forecasting economic activity with targeted predictors", *International Journal of Forecasting*, 31(1): 188-206.

Caggiano, G., Kapetanios, G. and Labhard, V., (2011), "Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK", *Journal of Forecasting*, 30(8): 736-752.

Diebold, F. X. and Mariano, R. S. (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, 13(3): 134-144.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., (2004), "Least angle regression", *Annals of Statistics*, 32(2): 407-499.

Eurostat (2013), *Handbook on quarterly national accounts.*

Foroni, C., and Marcellino, M., (2014), "A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates", *International Journal of Forecasting*, 30(3): 554-568.

Giannone, D., Reichlin, L. and Small, D., (2008), "Nowcasting: The real-time informational content of macroeconomic data", *Journal of Monetary Economics*, 55(4): 665-676.

Girardi, A., Golinelli, R., and Pappalardo, C., (2017), "The role of indicator selection in nowcasting euro-area GDP in pseudo-real time", *Empirical Economics*, 53(1): 79-99.

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

Kim, H. H., and Swanson, N. R., (2014), "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence", *Journal of Econometrics*, 178: 352-367.

Kim, H. H. and Swanson, N. R., (2018), "Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods", *International Journal of Forecasting*, 34(2): 339-354.

Kitchen, J., and Monaco, R., (2003), "Real-time forecasting in practice: The U.S. treasury staff's real-time GDP forecast system", *Business Economics*, 4(38): 10-19.

Jokubaitis, S., Celov, D. and Leipus, R., (2021), "Sparse structures with LASSO through principal components: Forecasting GDP components in the short-run", *International Journal of Forecasting*, 37(2): 759-776.

Panagiotelis, A., Athanasopoulos, G., Hyndman, R. J., Jiang, B., and Vahid, F., (2019), "Macroeconomic forecasting for Australia using a large number of predictors", *International Journal of Forecasting*, 35(2): 616-633.

Pashourtidou, N., Papamichael, C., and Karagiannakis, C., (2018), "Forecasting economic activity in sectors of the Cypriot economy", *Cyprus Economic Policy Review*, 12(2): 24-66.

Schumacher, C., (2010), "Factor forecasting using international targeted predictors: The case of German GDP", *Economics Letters*, 107(2): 95-98.

Stock, J. H. and Watson, M. W., (2002), "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association*, 97(460): 1167-1179.

Stock, J. H. and Watson, M. W., (2004), "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting*, 23:405-430.

Stock, J. H. and Watson, M. W., (2006), "Forecasting with many predictors", *Handbook of Economic Forecasting*, 1: 515-554.