Forecasting exports with targeted predictors

Francisco Dias
Banco de Portugal

Nuno Lourenço
Banco de Portugal and Nova SBE

António Rua
Banco de Portugal and Nova SBE

Abstract
This article applies factor models to forecast monthly Portuguese exports by resorting to an international dataset covering the country's main trading partners. We find noteworthy forecasting gains up to 12-month ahead when soft indicators for these countries are pooled and predictors are pre-selected prior to factor estimation. Resorting solely on national data and with no pre-selection of predictors yields greater forecasting accuracy when nowcasting. Hence, data from Portugal's main trading partners is more informative to produce h-step ahead forecasts. In turn, when hard and soft data are pooled, forecast accuracy is, in general, not enhanced. (JEL: C38, C55, F47)

Introduction

Forecasting macroeconomic time series is of utmost importance for fiscal and monetary policymakers to monitor or assess developments in any economy. Recent advances on short-term forecasting have drawn on the use of large datasets, where progress in information technology allows nowadays to access and handle hundreds of economic time series in real-time. Hard and soft data are at the core of this data-rich environment. While the former are based on quantitative information, the latter builds on surveys of economic activity that are characterized by the qualitative nature of their questions (e.g. regular harmonised surveys conducted by the European Commission for different sectors in the European Union). The interest in relying on soft data to forecast macroeconomic variables has been emphasized in the literature (see, for instance, Bańbura and Rünstler (2011) and Hansson *et al.* (2005) for an application to forecast GDP growth). A key advantage of qualitative indicators lies on their timeliness, as most surveys are released a few days after the reference period. Their high signal-to-noise ratio provides substantial informational content on the state of the economy and their encompassing nature allows for a wide sectoral coverage. Furthermore, since some questions concern future developments, they provide early information on the possible evolution of the economy. As these soft data series are not subject to revisions, real-time reliability is assured.

E-mail: fadias@bportugal.pt; nalourenco@bportugal.pt; arua@bportugal.pt

The use of soft data in addition to hard data for short-term forecasting has been notably attractive in Europe, where surveys are widely available for a long time (see, for example, Schumacher (2007) for an application to forecast GDP for Germany, Rünstler *et al.* (2009) to forecast GDP for several European countries and Angelini *et al.* (2011) for the euro area). While exploiting information from a data-rich environment has been widely documented in the literature, few authors have focused on the explicit role of adding foreign data to forecast national macroeconomic variables. In this direction, one should highlight the contributions by Brisson *et al.* (2003), where they take on board the predictive content of the United States variables as well as from other countries' to forecast Canadian real economic activity and inflation. Likewise, Schumacher (2009) considers the role of euro area and remaining G7 countries to forecast German GDP.

In a data-rich environment, forecasting macroeconomic variables amounts to extracting useful information from a large number of predictors. Factor models have been quite popular for such exercises, where the informational content from a large panel of time series is summarized in a few number of factors which are then used for forecasting purposes. Amongst the applications on the use of factor models stand out the seminal contributions by Stock and Watson (1999, 2002a,b) to forecast major macroeconomic variables for the United States, Marcellino *et al.* (2003) for euro-wide inflation and real activity and Rünstler *et al.* (2009) for a cross-country study comprising several European countries.

However, the use of additional data for forecasting with factor models might not improve forecast accuracy. In fact, extending a dataset for factor estimation can lead to worse forecasting results if the additional series are noisy or if forecasting power is provided by a factor that is dominant in a smaller dataset but turns out to be a dominated factor in a larger dataset (see Boivin and Ng (2006)). Reducing the influence of uninformative predictors to forecast a macroeconomic variable has given rise to a new stream of literature. In this regard, Bai and Ng (2008) proposed penalized regression techniques to target predictors, in particular, Least-Angle Regression with Elastic Net (henceforth LARS-EN), where selection of a subset of predictors prior to factor estimation is conducted to forecast US inflation. To further illustrate the importance of screening predictors prior to factor estimation, Schumacher (2007) provides an application to forecast German GDP and Li and Chen (2014) focus on several important variables in tracking the economy and monetary policies in the United States.

We investigate the role of information contained in foreign data to forecast international trade flows, with emphasis on exports of goods for Portugal, by extending the dataset to cover the country's main trading partners.[1] Thus, we contribute to the strand of the literature that relies on international data to forecast national variables. The high degree of interrelatedness of the Portuguese economy with the rest of the world lies at the core of this heterogeneous dataset. Given its large size, we then use LARS-EN pre-selection of predictors and infer on the usefulness of selection of predictors prior to factor estimation to enhance forecast performance. We exploit timely monthly data to nowcast and forecast Portugal's exports of goods on a monthly basis up to 12-month ahead.

The article proceeds as follows. Section 2 provides a quick glance at Portuguese exports. Section 3 introduces the analytical framework used in the forecasting exercise. Section 4 describes the underlying dataset used in the empirical application. The empirical results are discussed in Section 5 and Section 6 concludes.

A quick glance at Portuguese exports

Portugal has made notable progress in increasing its integration in world trade in the last few decades, with the accession to the European Economic Community in 1986 showing a particular leap forward. The relative importance of exports in the economy has grown gradually. However, it declined sharply in 2009 with the collapse of world trade, having gained pace thereafter (Figure 1).

The country's share of exports of goods in GDP in nominal terms has increased around 11 percentage points since 1993, standing roughly at 27 per cent in 2016. In the aftermath of the recent economic and financial crisis, Portugal experienced a gradual reallocation of inputs towards the production of goods for foreign markets. As a result, the relative importance of exports of goods increased markedly following the Great Recession.

A cross-country comparison with the euro area initial member states plus Greece (EA-12) portrays that Portugal stands as one of the countries that experienced the sharpest increase in the share of exports of goods in GDP during the period under study (Figure 2). It ranks ahead of the majority of the EA-12 member states, being surpassed only by The Netherlands and Germany.

1. Previous work on forecasting Portuguese exports of goods include Cardoso and Duarte (2006), who focus on forecasting nominal exports of goods using a small number of soft indicators through a bridge model.

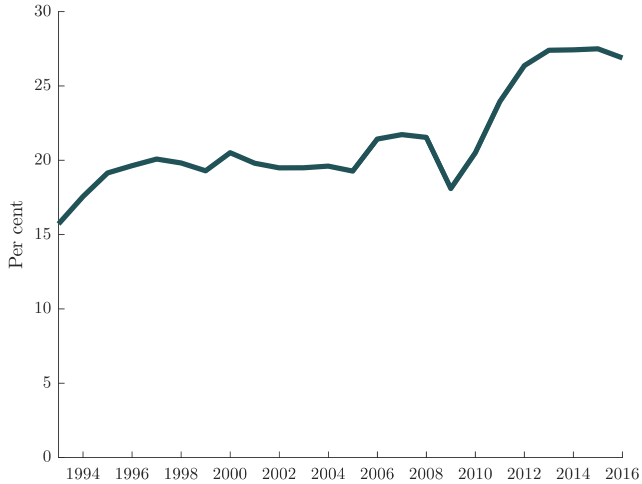

FIGURE 1: Share of exports of goods in GDP in nominal terms.

Source: Statistics Portugal.

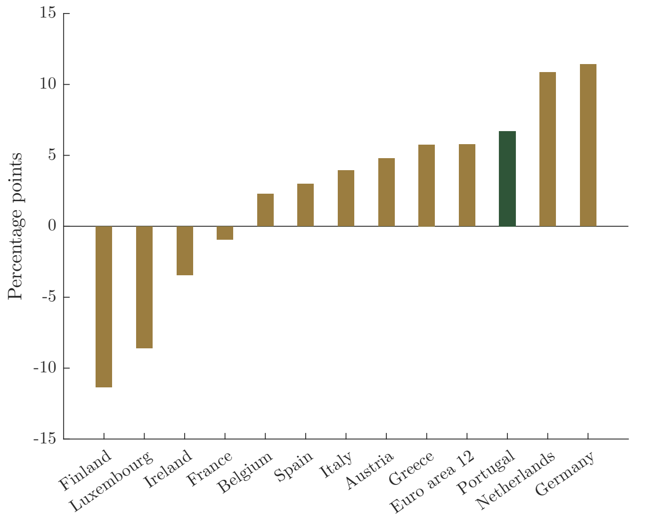

FIGURE 2: Change in the share of exports of goods in GDP in nominal terms between 2000 and 2016.

Source: Eurostat.

Concerning the main destinations of Portugal's exports of goods, euro area countries account for a large fraction. In Table 1, the main trading partners in 2016 are listed. In particular, exports to Spain comprise more than one fourth of total exports of goods, whereas France and Germany account for more than 10 per cent.

Main trading partners	**Shares (in per cent)**
Spain	26.2
France	12.6
Germany	11.6
United Kingdom	7.0
United States	4.9
Netherlands	3.7
Italy	3.4
Angola	3.0
Belgium	2.4

TABLE 1. Main destinations of Portuguese exports of goods in 2016.

Econometric framework

Factor models

We begin with a discussion of the factor model representation that motivates forecasting in a data-rich environment. Let X_t be an N-dimensional column vector of time series of predictor variables, observed for $t = 1, ..., T$. The aim of the exercise lies in representing these variables with a factor model representation and using the estimated factors to derive h-step ahead forecasts of the variable of interest, y, that is, y_{t+h}, where h denotes the forecast horizon. The variables in X_t are represented as the sum of two orthogonal components: the common component, driven by a small number of unobserved common factors that accounts for most of the co-movement among the variables; and the idiosyncratic component, driven by variable-specific shocks.

The data generating process for X_t admits a static factor representation written as:

$$X_t = \Lambda F_t + \xi_t \tag{1}$$

where $F_t = (f_{1t}, ..., f_{rt})'$ is an $(r \times 1)$ vector of non-observable factors, Λ is an $(N \times r)$ matrix of unknown factor loadings and ξ_t denotes an N-dimensional vector of idiosyncratic terms. As pointed out by Stock and Watson (2002b), unobserved factors can be estimated consistently through principal components under fairly general assumptions.

Factor estimation by principal components aims at maximizing the explained variance in the whole dataset. Typically, the first few top-ranked principal components capture a sizeable fraction of the co-movement among the series in the dataset. Once the number of factors has been selected, the variable to be forecast is projected eventually on its lags and on the set of r estimated factors. This yields the following forecasting equation for the variable of interest:

$$y_{t+h} = \alpha_0 + \sum_{i=1}^{r} \alpha_i \hat{F}_{t,i} + \sum_{j=0}^{p} \delta_j y_{t-j} + \varepsilon_{t+h} \tag{2}$$

where α_0 is a constant term, α_i denotes the coefficients pertaining to $\hat{F}_{t,i}$, i.e., the principal component estimates of the factors in Equation (1), y_{t-j} accounts for the autoregressive component of the regression, where δ_j are the corresponding coefficients and p the number of lags.

The LARS-EN algorithm

Typically, when factors are estimated, the informational content other than the one conveyed by the small set of r factors is ignored, thus, it can disregard useful information for the variable to be forecast or the forecast horizon at stake. Following Bai and Ng (2008), forecasting using targeted predictors is considered. In other words, the relationship between y_{t+h} and X_t is analyzed in order to select a subset of predictors $X_{t,\mathcal{A}} \subseteq X_t$ prior to factor estimation.

We now describe a method based on penalized regressions that performs subset selection and shrinkage by dropping uninformative regressors. Put differently, the regression coefficients of those variables that are less informative for predicting the targeted variable are penalized. Following Zou and Hastie (2005), Bai and Ng (2008) suggest the use of the EN optimization problem which is given by:

$$\underset{\beta}{min} \left\{ RSS + \lambda_1 \sum_{j=1}^{N} |\beta_j| + \lambda_2 \sum_{j=1}^{N} \beta_j^2 \right\} \tag{3}$$

where RSS is the residual sum of squares from a regression of y_{t+h} on all available regressors, and λ_1 and λ_2 penalize with the L_1- and L_2-norm of β, respectively.

The L_1 penalty solves

$$\underset{\beta}{min} \left\{ RSS + \lambda_1 \sum_{j=1}^{N} |\beta_j| \right\} \tag{4}$$

where the tuning parameter λ_1 controls for the amount of shrinkage, and thus for the number of parameters that are set to zero. The method adds λ_1 regularization to ordinary least squares regression, yielding solutions that are sparse in terms of the regression coefficients. This is also know as the Least Absolute Shrinkage and Selection Operator (LASSO) solution-type of Tibshirani (1996).

In turn, the L_2 penalty solves

$$\underset{\beta}{min}\left\{RSS+\lambda_2\sum_{j=1}^{N}\beta_j^2\right\} \tag{5}$$

which for $0 \leq \lambda_2 < \infty$ shrinks toward zero the coefficients of the uninformative predictors. This is also know as the L_2 penalty of ridge regression.

By combining both penalties, i.e., the virtues of LASSO and ridge regression, the EN in Equation (3) allows for shrinkage of coefficients, elimination of regressors and efficient selection of variables within the dataset.

LARS provides an efficient algorithm to solve the EN minimization problem (see Zou and Hastie (2005)). The LARS algorithm estimates β and selects the subset of predictors $X_{t,\mathcal{A}} \subseteq X_t$ by solving the optimization criterion in Equation (3), given the parameters λ_1 and λ_2. In practice, the calibration of λ_1 is recast as a rule for the maximum number of variables with non-zero β_j included in the analysis, i.e., the number of regressors $N_{\mathcal{A}} \leq N$ to be included in $X_{t,\mathcal{A}}$. The procedure works as follows. It starts with all coefficients equal to zero and finds the most correlated predictor with the variable to be forecast. It takes the largest step possible in the direction of this predictor until a second predictor has as much correlation with the current residual. Instead of continuing along the first predictor, LARS proceeds in a equiangular direction between the two predictors until a third variable earns its way into the most correlated set. LARS then proceeds equiangularly between the three predictors, that is, along the least angle direction, until a fourth variable enters and so on. The algorithm builds up estimates $\hat{\mu} = X\hat{\beta}$ in successive steps, each step adding one covariate to the model, so that after k steps just k of the $\hat{\beta}_j$'s are non-zero (see Efron *et al.* (2004) for details).

Data

The forecasting exercise comprises forecasting the growth rate of a key macroeconomic variable, nominal exports of goods for Portugal. This variable is released on a monthly basis 40 days after the reference period by Statistics Portugal without any seasonal or calendar adjustment.

We focus on forecasting the year-on-year growth rate of the series. Besides allowing to tackle deterministic seasonality, this choice can be motivated by several other reasons, such as the high volatility underlying month-on-month growth rates of nominal trade data or the larger resemblance between variables measured in year-on-year terms and the profile of several qualitative indicators. By considering year-on-year rates of change, noise in the data is reduced and data irregularities are smoothed out. For further discussion, see Esteves and Rua (2012). However, even when modelling the dependent variable as a year-on-year growth rate, calendar effects or moving holidays can be sizeable and are expected to impact the outcome variable. These effects are addressed resorting to deterministic variables to be described in the forecasting exercise.

Data for Portugal builds on a monthly dataset described in detail in Dias *et al.* (2015, 2018) comprising business and consumer surveys, retail sales, industrial production, turnover in industry and services, employment, hours worked and wage indices in industry and services, overnight stays in Portugal, car sales, cement sales, vacancies and registered unemployment, energy consumption, imports of goods, real effective exchange rate, Portuguese stock market index and ATM/POS series. Furthermore, we extend this dataset to include disaggregated data on consumer and producer prices.

Besides national data, we augment the dataset with international monthly data to cover information for Portugal's main trading partners. These include Spain, France, Germany, United Kingdom, United States, Netherlands, Italy and Belgium. Data for Angola are scarce, hence this country was disregarded from the dataset. The monthly dataset spans january-2000 to december-2016.

For each trading partner, the panel of variables includes the main quantitative measures of economic activity, as well as qualitative assessments amounting, on average, to 80 series per country and to 766 series overall. The series were selected to represent broadly business and consumer surveys of economic activity, prices, retail trade, manufacturing and services and labour market. The splitting of the number of variables into hard and soft data is provided in Table 2.[2]

	Number of series	**Soft data**	**Hard data**
Portugal	145	39	106
Spain	82	41	41
France	81	40	41
Germany	80	39	41
United Kingdom	80	39	41
United States	75	20	55
The Netherlands	77	39	38
Italy	80	39	41
Belgium	66	41	25
Total	766	337	429

TABLE 2. Composition of the dataset.

2. A list of all series and data sources is available from the authors upon request.

In the case of Portugal and Spain, for a limited number of series we resort to the Expectation-Maximization algorithm suggested by Stock and Watson (2002a) to balance the dataset at the beginning of the sample period, since some series were available over a shorter time span. In general, with the exception of survey data, logarithms were taken for all non-negative series that were not already in rates or percentage units. Most series were differenced to achieve stationarity. Additionally, the series were further screened for outliers, where the adjustment corresponded to replace observations of the transformed series with absolute deviations exceeding six times the interquartile range by the median value of the preceding five observations, following Stock and Watson (2005).

Forecasting exercise

We begin this section with a detailed description of the design of the forecasting exercise. This entails fully recursive parameter estimation and factor estimation after the selection of the targeted predictors using the LARS-EN algorithm. Thus, we do not restrict the set of targeted predictors to be the same for each time period. Instead, predictors are selected at each point in time for each horizon and the forecasting equation is re-estimated after the new factors are estimated. We also consider the case where no pre-selection of predictors is applied, i.e., using the standard factor model approach.

As the benchmark model, we consider the usual AR(p) with the number p of autoregressive terms determined by the standard BIC criterion. We augment this model to account for calendar effects with three deterministic variables: the number of working days in each month and two dummy variables for the two moving holidays, Easter and Carnival.

An out-of-sample exercise is conducted to assess the relative performance of the factor model with targeted predictors against the benchmark. The number of estimated factors to be included in the forecasting equation is determined by minimizing a modified version of the BIC criterion suggested by Stock and Watson (1998).

The out-of-sample period spans january-2009 to december-2016, corresponding to half of the available sample period and the forecasting exercise is based on rolling window estimation with a window size equal to 96 monthly observations (8 years), which coincides with the typical average length of the business cycle. Rolling window estimation enhances model flexibility and time-varying parameters to cope with potential varying predictive content of the dataset. All the potential predictors are available for time t when exports of goods are also known. However, in the case of soft data, when exports for $t-1$ are released, data for period t are known. Hence, when considering only soft data, one can consider nowcasting besides forecasting from 1 to 12-month ahead.

Model performance is evaluated using the Mean-Squared Forecast Error (MSFE) and we compute the relative MSFE using the augmented autoregressive model as the benchmark. Hence, a ratio lower than unity means that the competing model outperforms the benchmark. We evaluate to what extent the forecasting accuracy gains are statistically significant through the Clark and West (2007) test procedure.

In the empirical analysis to follow, we examine two alternative panels, where different datasets are considered. First, we analyze soft data driven forecasts. Thus, we exploit survey-based indicators for Portugal and for its main trading partners. Secondly, we extend the dataset, so that hard and soft data for the countries are pooled for the forecasting exercise.

Soft data driven forecasts

Table 3 reports the forecasting results with soft-based datasets with targeted predictors, i.e., with LARS-EN pre-selection setting $\lambda_2 = 0.25$ as in Bai and Ng (2008) and Schumacher (2009), and with no pre-selection of predictors. In case pre-selection of predictors has been applied, the number of chosen predictors is discretized in each row, $N_{\mathcal{A}} = \{30, 40, ..., 150\}$. Entries in the bottom of the table correspond to the case where no pre-selection is applied. The underlying datasets comprise only soft indicators amounting to 337 series when considering all countries and 39 series when resorting only on national data. Entries in the table refer to the relative MSFEs of the factor model *vis-à-vis* the augmented univariate autoregressive forecast for different forecast horizons. Shaded entries denote the minimum relative MSFE for each forecast horizon.

	Forecast horizon												
	0	1	2	3	4	5	6	7	8	9	10	11	12
Targeted predictors ($N_{\mathcal{A}}$)													
30	1.12	1.06	0.89	0.81	0.83	0.90	0.89	1.04	0.81	0.90	1.23	1.04	1.17
40	1.15	1.03	0.71	0.73	0.71	1.06	0.66	0.91	0.85	0.93	0.89	0.71	1.01
50	0.95	0.96	0.64	0.64	0.73	1.11	0.53	0.81	0.90	0.83	0.76	0.69	1.17
60	0.92	0.83	0.63	0.70	0.73	1.08	0.49	0.70	0.83	0.77	0.81	0.65	1.04
70	0.91	0.84	0.63	0.66	0.73	0.97	0.50	0.69	0.92	0.67	0.85	0.53	0.90
80	0.94	0.87	0.66	0.69	0.76	1.17	0.51	0.77	0.97	0.76	0.96	0.59	0.82
90	0.87	0.88	0.68	0.72	0.78	0.91	0.56	0.83	1.10	0.82	0.97	0.63	0.97
100	0.93	0.94	0.65	0.77	0.65	0.81	0.69	0.81	1.09	0.96	0.88	0.66	0.99
110	0.96	0.84	0.69	0.75	0.51	0.91	0.63	0.82	1.03	0.98	0.79	0.69	0.99
120	0.89	0.94	0.70	0.73	0.50	0.94	0.72	0.96	0.95	1.04	0.83	1.03	0.92
130	0.95	0.92	0.74	0.70	0.52	1.07	0.84	1.09	0.97	0.88	0.86	1.15	0.81
140	0.96	0.89	0.73	0.71	0.52	1.10	0.93	0.96	1.06	0.97	1.03	1.21	0.87
150	0.91	0.85	0.72	0.71	0.58	1.14	1.02	0.90	1.05	0.98	1.13	1.35	0.93
No pre-selection													
All series	0.89	0.82	0.71	0.68	0.83	1.39	1.60	1.85	2.19	1.39	1.43	1.54	1.42
PT series only	0.78	0.86	0.87	0.78	0.84	0.97	0.96	1.07	0.94	1.07	1.05	2.14	2.45

TABLE 3. Relative MSFE of soft data driven forecasts *vis-à-vis* the benchmark.

A quick overview of the results reveals most of the entries are below one, showing that there are, in general, forecasting gains using factor models *vis-à-vis* the benchmark.

When nowcasting, greater forecasting gains are achieved by using national soft data only and with no pre-selection of predictors, and these exceed 20 per cent. This may reflect that data from trading partners convey more informational content for forecasting purposes. In this regard, forecasting accuracy gains are noteworthy when soft data for Portugal's trading partners are exploited and these gains are further enhanced if LARS-EN pre-selection of predictors is applied. For forecasting 1-month ahead, although the maximum gain is near 20 per cent attained with all series, i.e., without pre-selection, a similar figure can be delivered considering only 60 targeted predictors. For the forecast horizons from 2- up to 12-month ahead, the use of targeted predictors is consistently a dominant strategy delivering gains ranging from almost 20 per cent up to around 50 per cent *vis-à-vis* the benchmark. In general, the best forecast performance is achieved with no more than 70 variables chosen out of 337 and the forecasting accuracy gains are statistically significant over the forecast horizons.

To shed some light on the composition of the set of targeted predictors used for factor estimation, Figures 3 and 4 provide plots for the average share of targeted predictors from each country and sectoral survey, respectively, for different number of predictors (N_A) and forecast horizons (h). By looking at Figure 3, one can see that the average share of selected series from Portugal increases with both the forecast horizon and the number of selected predictors, going from less than 5 per cent to more than 15 per cent. Focusing on the most important trading partners, the same holds broadly for Spain. In the case of France, the average share of series decreases with the forecast horizon and with the number of predictors, going from around 30 per cent to less than 10 per cent. In turn, for Germany, the average share is particularly important for shorter horizons (around 15 per cent) and less relevant for horizons close to one year, being relatively stable across the number of selected predictors.

In Figure 4, one can see that the manufacturing survey variables are very relevant to forecast at shorter horizons, attaining around 50 per cent for a small number of predictors, and its importance decreases with the forecast horizon and number of predictors. In contrast, when considering consumers' survey, the share increases with the forecast horizon and to a less extent with the number of predictors, reaching around 40 per cent. Shares in the remaining surveys are relatively stable, with services representing around 10 per cent and retail trade and building around 15 per cent.

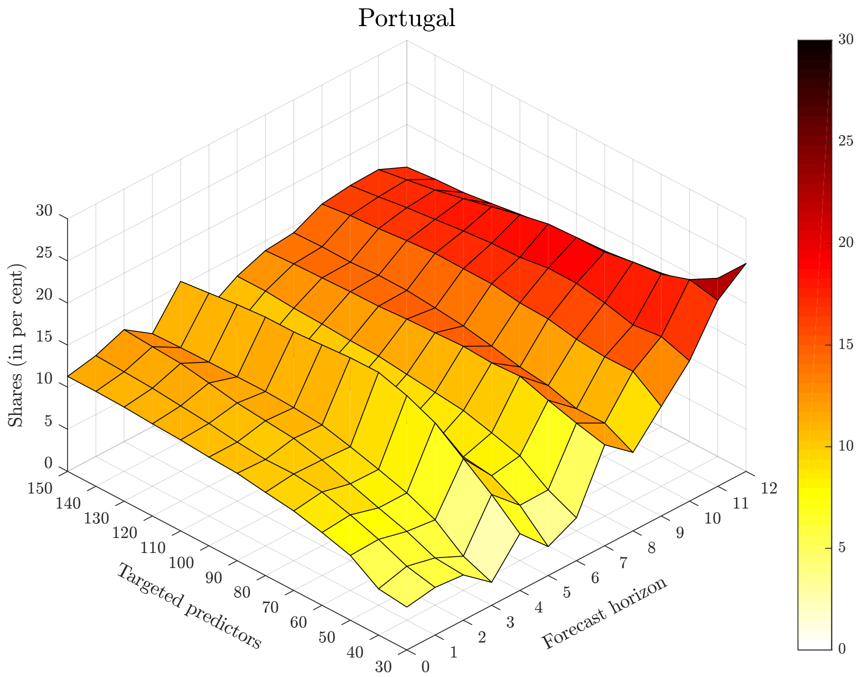

FIGURE 3: Average share of targeted predictors from each country for different number of predictors and forecast horizons.

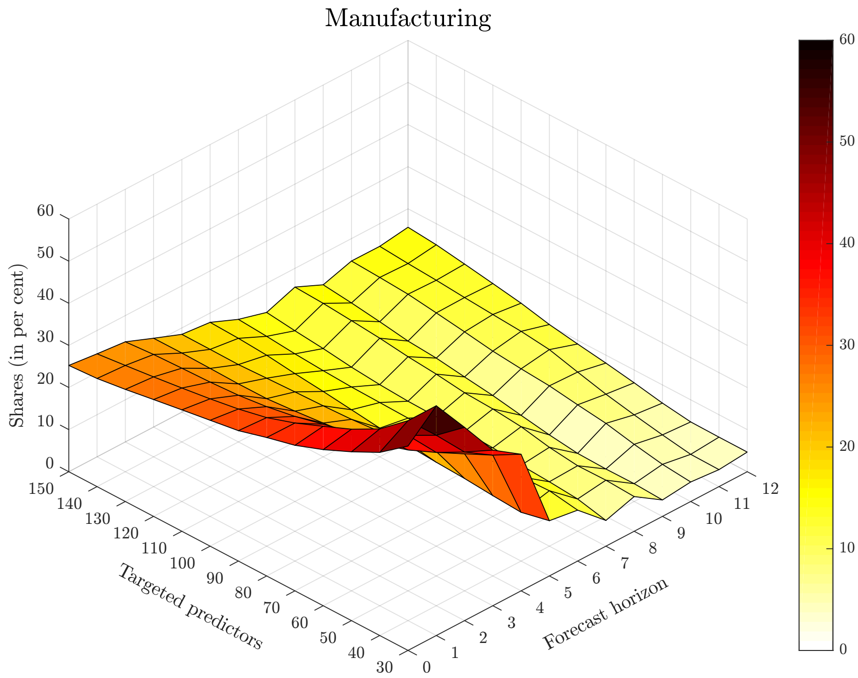

FIGURE 4: Average share of targeted predictors from each survey for different number of predictors and forecast horizons.

The EN algorithm involves the choice of λ_2, which penalizes with the L_2-norm of β. As in Bai and Ng (2008), we considered $\lambda_2 = \{0.5, 1.5\}$ as a robustness check. We find that the results are not very sensitive to this choice, which is in line with the findings by Bai and Ng (2008) and Schumacher (2009).

Pooling hard and soft data

We now extend the analysis where hard and soft data are pooled. As such, we exploit hard indicators for Portugal and its main trading partners, besides the previouly used soft-based dataset. The results are reported in Table 4. A quick glance at the results shows that the inclusion of hard data does not seem to bring additional predictive power for the longer horizons *vis-à-vis* the benchmark. In turn, the forecasting gains at shorter horizons are, in general, similar to those obtained when one resorts solely on soft data. One should note that LARS-EN pre-selection of predictors enhances forecast accuracy in comparison to no pre-selection and seems to play a role when forecasting at shorter horizons.

	Forecast horizon											
	1	2	3	4	5	6	7	8	9	10	11	12
Targeted predictors ($N_{\mathcal{A}}$)												
30	1.12	0.90	1.17	0.88	1.25	1.31	1.33	1.50	0.94	1.18	1.82	1.16
40	1.10	0.92	1.12	0.76	1.18	0.84	1.18	1.50	0.95	1.22	1.12	1.15
50	0.98	0.87	1.10	0.77	1.17	0.82	1.27	1.05	1.36	1.34	1.36	1.13
60	0.98	0.77	0.99	0.76	1.16	0.95	0.78	1.24	1.20	1.33	1.44	1.25
70	0.84	0.67	0.98	0.72	1.52	0.87	0.92	1.30	1.24	1.56	1.37	1.27
80	0.91	0.56	1.05	0.64	1.47	0.89	0.86	1.40	1.32	1.67	1.37	1.42
90	0.83	0.57	1.06	0.58	1.60	0.96	0.89	1.44	1.45	1.88	1.57	1.48
100	0.98	0.53	1.15	0.53	1.68	1.08	1.04	1.32	1.28	1.64	1.69	1.61
110	0.86	0.61	1.14	0.49	1.49	1.26	1.01	1.42	1.48	1.67	1.62	1.50
120	0.98	0.69	1.00	0.57	1.40	1.30	1.19	1.56	1.44	1.63	1.68	1.61
130	0.97	0.68	1.01	0.62	1.37	1.46	1.30	1.67	1.39	1.75	1.76	1.72
140	0.90	0.74	1.12	0.64	1.36	1.58	1.44	1.62	1.43	1.87	1.97	2.01
150	0.92	0.72	1.07	0.66	1.47	1.69	1.46	1.80	1.72	2.08	2.05	1.97
No pre-selection												
All series	0.86	0.82	1.18	1.46	2.26	2.43	2.73	3.78	4.06	3.57	4.28	4.64
PT series only	0.92	0.87	1.15	1.52	1.96	1.40	2.57	2.53	2.10	3.22	2.00	1.46

TABLE 4. Relative MSFE of soft and hard data driven forecasts *vis-à-vis* the benchmark.

Concluding remarks

This article exploits the role of international datasets for forecasting in a data-rich environment the Portuguese exports of goods on a monthly basis. Drawing on the informational content of the country's main trading partners, we document noteworthy forecasting gains up to 12-month ahead when soft indicators for these countries are pooled and predictors are pre-selected prior to factor estimation through the LARS-EN algorithm. In general, the best forecast performance is achieved with no more than 70 variables chosen. We find that forecasting accuracy gains delivered by factor forecasts using targeted predictors are statistically significant. Moreover, pooling hard data with soft data does not seem to bring additional predictive power for forecasting exports of goods.

References

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). "Short-term forecasts of euro area GDP growth." *Econometrics Journal*, 14, C25–C44.

Bańbura, M. and G. Rünstler (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, 333–346.

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

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

Brisson, M., B. Campbell, and J. W. Galbraith (2003). "Forecasting Some Low-predictability Time Series Using Diffusion Indices." *Journal of Forecasting*, 22, 515–531.

Cardoso, F. and C. Duarte (2006). "The Use of Qualitative Information for Forecasting Exports." *Banco de Portugal, Economic Bulletin, Winter 2006*, pp. 67–74.

Clark, T. E. and K. D. West (2007). "Approximately normal tests for equal predictive accuracy in nested models." *Journal of Econometrics*, 138, 291–311.

Dias, F., M. Pinheiro, and A. Rua (2015). "Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence." *Economic Modelling*, 44, 266–272.

Dias, F., M. Pinheiro, and A. Rua (2018). "A bottom-up approach for forecasting GDP in a data-rich environment." *Applied Economics Letters*, 25, 718–723.

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

Esteves, P. S. and A. Rua (2012). "Short-Term Forecasting For The Portuguese Economy: A Methodological Overview." *Banco de Portugal, Economic Bulletin, Autumn 2012*, pp. 137–149.

Hansson, J., P. Jansson, and M. Löf (2005). "Business survey data: Do they help in forecasting GDP growth?" *International Journal of Forecasting*, 21, 377–389.

Li, J. and W. Chen (2014). "Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models." *International Journal of Forecasting*, 30, 996–1015.

Marcellino, M., J. Stock, and M. Watson (2003). "Macroeconomic forecasting in the euro area: country specific versus euro wide information." *European Economic Review*, 47, 1–18.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze (2009). "Short-Term Forecasting of GDP Using Large Datasets: A Pseudo Real-Time Forecast Evaluation Exercise." *Journal of Forecasting*, 28, 595–611.

Schumacher, C. (2007). "Forecasting German GDP Using Alternative Factor Models Based on Large Datasets." *Journal of Forecasting*, 26, 271–302.

Schumacher, C. (2009). "Factor forecasting using international targeted predictors: The case of German GDP." *Economic Letters*, 107, 95–98.

Stock, J. and M. Watson (1998). "Diffusion indexes." *NBER Working Paper no. 6702*.

Stock, J. and M. Watson (1999). "Forecasting inflation." *Journal of Monetary Economics*, 44, 293–335.

Stock, J. and M. Watson (2002a). "Macroeconomic forecasting using diffusion indices." *Journal of Business and Economics Statistics*, 20, 147–162.

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

Stock, J. and M. Watson (2005). "Implications of dynamic factor models for VAR analysis." *NBER Working Paper Series 11467*.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the LASSO." *Journal of Royal Statistical Society Series B*, 58, 267–288.

Zou, H. and T. Hastie (2005). "Regularization and variable selection via the elastic net." *Journal of Royal Statistical Society Series B*, 67, 301–320.