UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 14-09
Statistics and Econometrics Series 05
March 2014

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

Selecting and Combining Experts from Survey Forecasts

Julieta Fuentes[1], Pilar Poncela[2], Julio Rodríguez[3]

Abstract

Combining multiple forecasts provides gains in prediction accuracy. Therefore, with the aim of finding an optimal weighting scheme, several combination techniques have been proposed in the forecasting literature. In this paper we propose the use of sparse partial least squares (SPLS) as a method to combine selected individual forecasts from economic surveys. SPLS chooses the forecasters with more predictive power about the target variable, discarding the panelists with redundant information. We employ the Survey of Professional Forecasters dataset to explore the performance of different methods for combining forecasts: average forecasts, trimmed mean, regression based methods and regularized methods also in regression. The results show that selecting and combining forecasts yields to improvements in forecasting accuracy compared to the hard to beat average of forecasters.

[1] Julieta Fuentes, Universidad Carlos III de Madrid and Banco Central de Reserva de El Salvador, Alameda Juan Pablo II, San Salvador, e-mail: julieta.fuentes@alumnos.uc3m.es; julieta.fuentes@bcr.gob.sv. Corresponding author
[2] Pilar Poncela, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049, Madrid, e-mail: pilar.poncela@uam.es
[3] Julio Rodríguez, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049, Madrid, e-mail:jr.puerta@uam.es

Introduction

The surveys of expert forecasts have proved to be relevant on the formation of macroeconomic expectations of economic agents. Although the forecast of the individual experts are made available by the corresponding institutions, the public usually focuses on a summary measure as the mean or median values of the forecasted variables, which generally are also provided. The empirical finding that simple forecast combinations, in particular the equal weights average, tend to outperform individual forecasts or more sophisticated schemes is one of the reasons that may explain this practice (Clemen, 1989; Makridakis et al., 1982; Stock and Watson, 2001 and Genre et al., 2013; among others). This is known as the "forecast combination puzzle" (see, for instance, Aiolfi et al., 2011).

The idea of examining combination methods that could exploit in a better way the information provided by multiple forecasts of the same variable has motivated a large literature on forecast combination (Bates and Granger, 1969; Marcellino, 2004; Timmermann, 2006; among others). The evidence has demonstrated that forecast combinations improve the performance of individual forecasts. Hence, recent research has focused on the optimal combination problem of the forecasts. There are several methods proposed in the literature to estimate combination weights. From the seminal paper by Bates and Granger (1969), to more recent sophisticated alternatives as, for instance, factor methods (Poncela et al., 2011) and Bayesian shrinkage combinations (Diebold and Pauly, 1990) there is an extense literature that has focused on this issue.

In general, to forecast the target variable y_{t+h} the previous methods combine the available set of predictors or individual forecasts $y^1_{t+h|t},\ldots,y^N_{t+h|t}$, assigning a positive weight to all of them. In this paper we are interested in performing a selection of informative experts from a survey, discarding some individual forecasts. Some forms of trimming have been proposed (Stock and Watson, 2004; Granger and Jeon, 2004 and Aiolfi and Favero, 2005) that in its simplest form is to discard α% of the lowest and highest values of the forecasts and then take the average of the remaing ones. When the number of forecasts to combine is high, the trimming technique can be quite aggressive (see, for instance, Samuels and Sekkel 2013, in the context of combination of forecast from models, instead of surveys). Our proposal is somewhat different, because it tries to remove the forecasts with redundant information or small predictive power about the target (not extreme one). Conflitti et al. (2012) approach for point forecast of the European Central Bank Survey of Professional Forecasters (ECB SPF) is the closest to our scheme. However, using the well-known Stock and Watson database, Fuentes et al. (2014) find that the selection of predictors achieved by the sparse PLS (SPLS) improves the forecast efficiency compared to the widely used competing models, including the least angle regression pure selection procedure.

Poncela et al. (2011) analyze several multivariate techniques to combine the expert forecasts. For each of the six U.S. Business Indicators from the Survey of Professional Forecasters (SPF from now on) for the period 1991-2008, they find that taking into account the target variable y_{t+h} in the process of dimension reduction, the combination outperforms the standard benchmark (simple average of individual forecasts) and, in particular, Partial Least Square (PLS) provides a good forcasting performance.

We explore the empirical performance of different ways to combine forecasts and to combine selected forecasts from surveys. We use the individual forecasts for U.S. economic variables, collected by the Philadelphia Federal Reserve Bank's SPF for the period 1991-2012. Also, we divide the entire sample into three periods to evaluate the robustness and the business cycle sensitivity of the results.

In particular, we propose to investigate the usefulness of SPLS, a technique that allows selecting and combining the informative predictors for a forecasted target.

The paper is organized as follows. Section 2 briefly presents the different combination and selection methods. Section 3 decribes some relevant features of the SPF dataset. Section 4 presents the empirical applications and the forecasting results. Finally, Section 5 concludes.

2. Forecast Combination and/or Selection Methods

In this section we briefly describe different techniques to combine the forecasts based on survey data. Some approaches include the information from the full panel of forecasters in the combination:

Ordinary Least Squares (OLS) and Partial Least Squares (PLS); while others perform a selection of forecasters before combining them: trimmed mean, Least Angle Regression (LARS) and SPLS.

2.1Trimmed mean

The trimmed mean is the mean computed by excluding a $\alpha/2\%$ of the lowest and highest values from a sample. For example, a mean trimmed by 50% has 25% of the largest forecast and 25% of the smallest forecasts removed.

This is a measure of central tendency and it is considered a robust estimator of location for a symmetric distribution, because it reduces the influence of outliers. The median can be considered as an extreme trimming method.

2.2 Partial Least Squares (PLS)

Partial Least Squares was developed by Wold (1966). PLS is a dimension reduction technique which constructs a scheme for extracting orthogonal latent factors based on the covariance between the predictors (X) and the dependent or forecasting variable (Y). In this particular case, $X_t = (y_{t+1|t}^1, \dots, y_{t+1|t}^N)$ is an N-dimensional vector of one step ahead forecasts of the target variable from the survey of panelists at time t, $\underset{t \times N}{X} = (\mathrm{X}_1^{'}\,\mathrm{X}_2^{'} \dots \mathrm{X}_t^{'})'$ and $Y = (Y_1\, Y_2 \dots Y_t)'$.

The PLS components are obtained iteratively. The first PLS factor $\hat{f}_{jt}^{PLS}$ can be computed from the eigenvalue decomposition of the matrix $M = X^{'}YY^{'}X$ .To find the second factor, we remove from the original data the part explained by the regression between the data and the first PLS component. Then, the eigenvalue decomposition is performed on the residual matrices from the previous iteration, which contains information that is orthogonal to the first component. The process continues until all factors have been extracted.

PLS has been applied for macroeconomic forecasting in Fuentes et al. (2014) and Kelly and Pruitt (2012).

It is important to note that Partial Least Squares focuses on maximizing the covariance between X and Y, in contrast to principal components regression, which focuses only on the variance (Hastie et al., 2008). Therefore, the linear combinations provided by PLS are oriented towards forecasting y_{t+1}.

2.3 Least Absolute Shrinkage and Selection Operator (LASSO)

Lasso is a selection and shrinkage method for regression. It was proposed by Tibshirani (1996) as a strategy to reduce the variance (sacrificing a little of bias) obtained by the OLS estimates. LASSO coefficients minimize a penalized residual sum of squares defined as follows:

$$\begin{aligned} &\hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \\ &\textit{subject to} \sum_{j=1}^{p} | \beta_j | \le \tau. \end{aligned} \tag{1}$$

The tuning parameter τ could be varied, if τ is sufficiently large then the constraint has no effect and the LASSO algorithm will yield the linear least squares estimates (OLS). But if τ is sufficiently small, the constraint tends to produce some zero coefficients enforcing sparsity in the solution.

LASSO does not have a closed form expression, but several efficient algorithms have been developed for solving it. LARS provides a numerical aproximation for performing LASSO. It is a stepwise variable algorithm that adds in each step one covariate to the model. At the first step it identifies the predictor most correlated with the response to form the active set of covariates. The β coefficient is moving towards its least squares coefficient until a covariate outside the active set has the same correlation with the residual. Once the second covariate in the active set has been selected, the process is repeated until all predictors have been entered in the model. We will retain the first k selected forecasts for their inclusion in an OLS regression.

2. 4 Sparse Partial Least Squares (SPLS)

Since PLS assigns weigths to all the predictors in a given data set, noise or irrelevant variables may affect the weights. To overcome this drawback, Chun and Keles (2010) propose a sparse PLS formulation (SPLS) in the context of biological data. The authors introduce an algorithm that searches at each step for informative variables by optimizing the following expression:

$$Min_{w,c} -\theta w' Mw + (1-\theta)(c-w)' M(c-w) + \lambda \, | \, c \, |_1 \qquad (2)$$

$$\textit{subject to } w' w = 1,$$

where λ encourages sparsity on the direction vector; c, is a surrogate direction vector being the sparse version of w and θ is a parameter which controls the effect of the concavity of the objective function and the closeness of the original direction vector (w) and the surrogate one (c).

The constraint leads to sparse linear combinations of the original predictors given in terms of the surrogate vector. The solution is achieved conducting OLS regression by using the sparse linear combinations given by the SPLS algorithm. Regularized PLS has been applied in economics (Fuentes et al., 2014 and Groen and Kapetanios, 2008).

2.5 Some intutitions

After introducing the forecasting combination methods, we are going to motivate them through the population case of combining N experts. This will allow us to understand when and how the different methods may work. To keep the expositition as simple as possible, we will present the case of just two linear combinations of one period ahead forecast $y^{i}_{t+1|t}$, i = 1, …, N.

Assume that since all experts want to forecast the same target value, they can be highly correlated. The OLS combination of the experts is given by:

$$y_{t+1} = \beta_0 + \beta_1 y^{1}_{t+1|t} + \ldots + \beta_N y^{N}_{t+1|t} + \varepsilon_{t+1}. \qquad (3)$$

Due to multicollinearity, the uncertainty in the estimation of the β's can be quite high. Therefore, using them for forecasting can produce unstable forecasts, specially in the turning points. For instance, assume that all experts have been quite collinear in the past but a few of them foresee the recession while the remaining ones do not.

The weigths and β's for each of the methods mentioned above are different. For the mean or average of forecasters $\beta_i = 1/N$, i=1, ... , N and $\beta_0 = 0$, while for the trimmed mean $\beta_i = 1/[(1-\alpha)N]$, if forecaster i is not trimmed and 0 otherwise, where 1-α is the percentage of central observations considered for the estimation of the mean, [(1-α)N] stands for the positive integer closest to (1-α)N and as in the previous case $\beta_0 = 0$.

Notice that with the average and the trimmed average forecasters we go from N individual forecasters to just 1 final combined forecast through

$$f_t = \sum_{i=1}^{N} \beta_i y_{t+1|t}^{i}. \tag{4}$$

In spite of its simplicity, the average of forecasters has been widely used and has been continuously analyzed (see, for instance, Clements and Harvey, 2009; Timmerman, 2006 and Genre et al., 2013; for some recent papers). As it is well known, the average of forecasts works well when the variance associate to each forecasters is the same (homogeneous forecasts), irrespective of the correlation among them.

In the case of heterogeneous forecasts, because some of the forecasts may be biased or even if we consider only unbiased forecast, the variance of each one of them may be different, the selection by the trimmed mean migth work well because it can discard biased forecasts and inefficient forecasters in the sense of high variance. A particular case that may be of interest for macroeconomics is the case of bimodal distribution of the cross section of forecasters. For instance, in the case of entering into a recession this might not been anticipated by all the forecasters, but only by some of them. In this case, the trimmed mean may not work as desired.

However, in cases were the forecast error from expert i is quite large, the trimmed mean might be a good solution. When the number of extperts is big, it can be a good method for choosing the central experts. In our forecasting problem, the experts were already selected by their continuity in participating in the survey and therefore our sample is already trimmed in some sense.

Equation (3) can be approximated by selecting some of the experts as LASSO does. Another alternative is to rely on a small set of linear combinations of the experts or factors that capture common behaviour of the experts.

Concerning the first approach, selecting some of the experts might yield better forecasting results if there is a group of "better" experts, that is, experts that systematically have a small MSE. If this is the case, it might be worth it selecting them and discarding the remaining ones.

As regards exploiting the common information shared by the experts, equation (3), seen as the combination rule for the experts, might be seen as the result of a two step procedure where first the experts are combined as follows

$$f_t^j = \sum_{i=1}^{N} \omega_{ij} y_{t+1|t}^i \,, \quad \text{j=1, …, k} \tag{5}$$

And then the k linear combinations of the forecast are regresses over the target

$$y_{t+1} = \gamma_0 + \sum_{j=1}^{k} \gamma_j f_t^j + \varepsilon_{t+1} \tag{6}$$

Notice that (6) should be estimated with information up to time t to generate true ex ante forecasts. Observe that (5) and (6) are equivalent to (3).

In cases were the experts are higly collinear, they may belong to a subspace of lower dimension than the number of experts. In this situation, finding as many components as the dimension of the subspace spanned by the experts might result in a lower RMSE. If these k linear combinations are oriented towards the forecasting target, the factors are obtained by PLS were in the simplest case of one common component $\omega_i \propto Cov(y_{t|t-1}^i, y_t)$ and $\gamma_1 = \dfrac{Cov[y_t, f_{t-1}^{pls}]}{V[f_{t-1}^{pls}]}$.

The flexibility of PLS and SPLS is that we can increase the number of components and cover cases of heterogeneous forecasters. As regards the case of PLS recall that the weights take into account the covariance with the target, giving more weight to those forecasters more correlated with the target. The first PLS component takes into account the univariante effect on the target of each forecaster. We seek for directions of high variance as well as high correlation with the target.

To give some intuitions where the PLS approach might work well, consider the case where the majority of forecasters are highly collinear but a few of them, for instance, a small group might foresee the recession although the majority do not. In this case the first PLS component gives weights to all forecasters as usual. The second PLS component, orthogonal to the first one, can capture the variation in the small group of forecasters that disagree if this variation is on the direction of the target. Notice that this effect cannot be captured by the average forecast.

However, there are cases were the weigths ω_i are not too high and their estimation uncertainty might not compensate their contribution to the final forecast. In this instance we can obtain linear combinations of the experts with a smaller MSE by setting some of the weights to zero.

Adittionally, there are other situations were sparsity might yield better results. In PLS we give some weight to all the experts. If there is a small group of them that foresee the recession, their weight is diluted by the number of experts, maybe not giving them enough weight. If we could introduce some sparsity by zeroing out some redundant experts or not very informative ones, the group that anticipates the turning point might be not too diluted and capture the change in the economy. We might need more than 1 sparse component to pick up the previous behaviour.

3. Empirical Application

To explore how selecting and combining expert forecasts works in practice, we employ the dataset generated by the Survey of Professional Forecasters (SPF). The SPF is conducted by the Federal Reserve Bank of Philadelphia since 1990, when overtook the American Statistical Association (ASA) and the National Bureau of Economic Research (NBER) project started in 1968.

The survey is conducted on a quarterly basis and asks the forecasters to provide projections for the next five quarters: the quarter in which the survey is conducted and subsequent four quarters. Advanced reports of government statistical agencies such as the U.S. Bureau of Economic Analysis (BEA) and the U.S. Bureau of Labor Statistics (BLS) of the previous quarter are reported in the survey questionnaiere sent to the panelists.

We use the one-step-ahead forecast of the consumer price index (CPI) and five variables of the U.S. Business Indicators group: Nominal Gross Domestic Product (NGDP), 3 Month Treasury Bill Rate (TBILL), AAA Corporate Bond Yield (BOND), Civilian Unemployment Rate (UNEMP) and Housing Starts (HOUSING). We assume that the logarithm of NGDP is integrated of order 1 and that the levels of the remaining series are difference stationary.

The considered panel of forecasts for all the variables starts in the first quarter of 1991 and ends in the fourth quarter of 2012. Because the dataset is particularly unbalanced, due to the entry and exit of forecasters, we apply two types of pre-treatments on the information from the SPF. First, as in Poncela et al. (2011) we reduce the pool of panelists by selecting those forecasters who met the following criteria: (i) they have been on the panel at least seven years and (ii) they have no more than four consecutive quarters without an active participation in the survey. Second, missing observations are filled with the marginal mean of the forecast up to the previous period given by the respondent. If the missing observations occur in the first period of the estimation sample, they are filled with the non revised or first estimated data of the previous period.

To perform a sensibility analysis and a robustness check, we divide the full sample in three periods. One forecasting sample includes only a period of economic expansion. On the contrary, the remaining two samples contain recession periods according to the NBER dating. Table 1 summarizes the estimation and forecasting samples.

The period of projection of the first sample spanning from the first quarter of 2000 to the fourth quarter 2003 contains the dot.com recession. In the third sample, the forecast sample covers from the first quarter 2007 to the fourth quarter 2012 including the last deep recession; while for the second sample is an expansionary period covering from the first quarter 2005 to the fourth quarter 2007.

Table 1
Estimation and forecasting subsamples

	Estimation subsample	Forecast subsample
M1	1991:02 to 1999:04-h	2000:01 to 2003:04
M2	1996:01 to 2004:04-h	2005:01 to 2007:04
M3	2001:01 to 2006:04-h	2007:01 to 2012:04

The entry and exit of panelists makes extremely difficult to have a long sample for estimating the competing methods. In this case it is not possible to perform an analysis for the full sample. The number of panelists considered oscillates between 14 and 19, depending on the variable and the subsample (see Table 2).

Table 2
Number of forecasters of each subsample

	BOND	CPI	HOUSING	NGDP	TBILL	UNEMP
00.1-03.4	14	14	14	16	15	16
0.5.1-07.4	16	16	17	18	16	18
07.1-12.4	14	17	15	18	16	19

4. Forecast results

To compare the forecast accuracy of the trimmed mean, OLS, PLS, LARS and SPLS combination schemes relative to the equal weighted benchmark, we use the relative root mean squared forecast error (RMSE):

$$\text{Re}\mathit{lative\ RMSE} = \frac{RMSE\ (method)}{RMSE\ (average)}. \tag{5}$$

An entry of less than one implies that the combination scheme outperforms the simple average forecast.

For the trimmed mean, we employ two values of α: 20% and 50%. As regards LARS, we will retain the first 5 and 10 selected forecasts for their inclusion in an OLS regression. The SPLS approach is implemented considering different values for the sparsity parameter (λ) = 0.2, 0.4, 0.6 and 0.8. The number of components considered for PLS and SPLS are k=1, 2 and 3.

4.1 Relative Performace of the Competing Methods

The forecasting results for the six variables and the three forecasting subsamples considered are shown in Tables 3 and 4. The comparisons performed suggest several notable features. First, they show that in all cases there is a method that works better than the simple average. We find predictive gains in techniques which perform a forecaster's selection: trimmed mean, LARS and SPLS. Second, SPLS yields the best forecasting performance in 72.2% of cases, and its accuracy is similar to the best alternative model in 16.7% of the remaining ones. Third, SPLS provides the best forecasting results for the variables BOND, CPI and UNEMP in all subsamples considered.

Table 3
Relative RMSE for BOND, CPI and UNEMP. h=1

BOND. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.016	1.003	1.132	0.971	1.011	1.100	**0.846**	0.980	1.018	0.873	1.013
0.5.1-07.4	0.921	0.887	0.786	0.909	0.740	0.795	0.727	**0.713**	0.784	0.827	0.850
07.1-12.4	0.936	0.963	0.968	0.928	0.937	0.935	**0.810**	0.814	0.934	0.833	0.940

CPI. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.011	1.023	1.044	0.962	0.939	1.049	**0.899**	0.939	1.014	0.911	0.987
0.5.1-07.4	0.986	1.021	1.236	0.869	1.005	1.217	**0.843**	1.005	1.130	0.979	1.221
07.1-12.4	0.962	0.955	1.293	0.869	0.654	0.855	0.832	**0.613**	0.815	0.775	0.870

UNEMPLOYMENT. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.973	0.889	1.934	0.840	1.124	1.413	**0.839**	1.006	1.402	1.067	1.272
05.1-08.4	0.932	0.858	1.285	0.697	0.953	1.082	**0.697**	0.838	0.889	0.919	1.133
07.1-12.4	0.940	0.847	1.288	0.766	0.900	0.921	**0.766**	0.876	0.900	0.969	0.992

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

Fourth, we find forecast improvements of SPLS with respect to the benchmark for TBILL. For the first subsample (00.1-03.4) SPLS provides the best performance and for the other two subsamples, in which LARS (10) gives the best results, it performs quite well and its accuracy is close to that of LARS. Fifth, there are two particular cases for which the benchmark has proven to be hard to beat: (i) the subsample (05.1-07.4) for NGDP and (ii) the subsample (07.1-12.4) for HOUSING, in both cases the trimmed mean (α=50%) is the only method able to produce accuracy gains over the simple average.

Table 4
Relative RMSE for TBILL, NGDP and HOUSING. h=1

TBILL. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.973	0.786	1.266	0.667	0.752	0.847	**0.650**	0.747	0.847	0.746	1.050
0.5.1-07.4	0.986	0.976	0.839	0.841	0.904	0.795	0.811	0.898	0.791	0.872	**0.780**
07.1-12.4	0.972	0.934	0.738	0.754	0.732	0.682	0.754	0.711	0.678	0.694	**0.655**

NGDP. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.006	1.005	2.275	**0.925**	1.128	1.318	0.926	1.135	1.301	1.406	1.565
0.5.1-07.4	1.006	**0.982**	1.785	1.306	1.370	1.751	1.235	1.365	1.719	1.601	1.655
07.1-12.4	1.005	1.015	1.548	1.040	1.010	1.077	1.035	**0.967**	1.077	0.994	1.172

HOUSING. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.996	0.980	0.824	0.841	0.777	0.823	**0.710**	0.758	0.823	0.773	0.837
0.5.1-07.4	0.952	0.967	0.796	0.906	0.742	0.807	0.774	0.700	**0.662**	0.724	0.750
07.1-12.4	0.951	**0.943**	2.136	1.549	1.460	1.681	1.407	1.463	1.657	1.581	1.743

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

Notice that in cases where the average of forecasters works well in comparison with OLS, 1 component (PLS or SPLS) should also work well. In this case selecting more components is not a wise solution since we will be approaching OLS. This can be seen, for instance, in the table for unemployment. The reason why PLS or SPLS might work better than the average of forecasts is that we can give more weight to those forecasters more correlated with the target. The key issue is not with extreme values (trimmed mean) but giving more weight to the most correlated forecasters with the target. On the contrary, if OLS outperforms the average of forecasts, then choosing more PLS or SPLS components should be advisable.

In general, the best performing SPLS models have just one component and a high degree of sparsity (λ=0.8). However, the number of panelists selected depends on the variable and on the period measured. For HOUSING and NGDP the number of informative forecasters is reduced significantly in all periods analyzed. For the rest of variables, the panel composition changes with the subsample.

4.2 Sensitivity Results to Business Cycle

4.2.1 RMSE

As can be seen from Tables 3 and 4, the relative performance of the competing models over all the subsamples and variables is quite similar. The inclusion of the financial crisis period of 2008-2009 in the third subsample does not seem to produce a notable effect in their forecasting performance with respect to the average of forecasters, except for HOUSING[4]. In this last case, the influence of the housing sector crash of 2006-2007 –which is considered the worst in the U.S history and the root of the globlal crisis- and its sluggish recovering, could explain the observed decline in the forecast accuracy of the different methods.

It is important to highlight that the size of the estimation sample varies among subsamples, as well as the forecasters involved in them; thus, the uncertainty of estimation also differs among them.

To evaluate in a more detailed way the forecasting results during the pre-crisis and crisis periods, the second subsample was extended until the fourth quarter of 2009. For this subsample, the estimation period starts in 1996:01 and ends at 2004:04 and the forecasting period covers from 2005:01 until 2009:04. Table 5 reports the relative RMSE results.

[4] In the case of BOND, it is observed deterioration in the accuracy of the forecast respect to the second sample.

The results for the extended sample 2 evidence that a number of the considered schemes outperform the equal weighted combination. In particular, with the exception of NGDP, the best performing method across variables is SPLS. The SPLS method improves on the benchmark and it yields gains between 16.7% and 37.1%; the greatest reduction in the RMSE was achieved by CPI. The good performance of some of the combination schemes for inflation has been found for the ECB SPF, over normal business cycle conditions (Genre et al., 2013). In the case of the Harmonised Index of Consumer Prices (HICP) for the Euro area, the authors link the positive results of the combination strategies to the correction of a persistent downward bias in the inflation forecasts.

Table 5
Relative RMSE. Forecasting Sample: 2005-2009

Variable	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
BOND	0.921	0.907	1.146	0.946	0.856	1.041	0.846	**0.833**	1.030	1.045	1.224
CPI	0.987	0.947	1.187	0.712	0.670	0.896	**0.629**	0.670	0.801	0.688	0.891
HOUSING	0.960	0.944	1.009	1.066	0.952	0.996	0.828	**0.804**	0.882	0.878	0.940
NGDP	1.000	**0.978**	1.697	1.055	1.147	1.492	1.043	1.063	1.425	1.241	1.705
TBILL	0.976	0.949	0.951	0.712	0.900	0.788	**0.708**	0.778	0.785	0.949	0.974
UNEMPL	0.935	0.952	1.166	0.680	0.894	0.979	**0.680**	0.877	0.974	0.992	0.952

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

Table 6 compares the RMSE of the extended subsample splitting the prediction errors in two groups; those related to the "Non-Crisis" period and a second group corresponding to the time period 2008:01-2009:02, defined as the recession by the NBER Business Cycle Dating Committee. For CPI, NGDP, TBILL and UNEMP, we find a relative better performance with respect to the average forecasts in the crisis period than in the non crisis period. BOND and HOUSING present the opposite behavior.

A stylized fact in the literature is that household investment leads the business cycle. As was mentioned before, the beginning of the crisis period of the housing sector is prior to that used in this comparison, even though still represents a deceleration period in which the results show a high sensitivity.

Table No.6
Relative RMSE for Non-Crisis and Crisis Period

Variable	SPLS	
	Non-crisis	Crisis
BOND	0.457	1.451
CPI	0.683	0.260
HOUSING	0.563	0.873
NGDP	1.757	0.849
TBILL	0.710	0.289
UNEMP	0.777	0.414

Source: Authors´ calculations. The table shows the ratio of RMSE of SPLS over the benchmark model.

During the recession period, the SPLS model squared forecast errors across variables reported an increase, reflecting the increased uncertainty associated with the financial crisis. For TBILL this phenomenon was observed in advance from the second quarter of 2007 to the third quarter of 2008. As previously noted, in the case of HOUSING, forecast errors also observed an increase during the year 2006 and for NGDP, BOND and CPI the forecast errors behaviour also seems to have changed in the year 2006 (see, Figure 1).

4.2.2 The Degree of Sparsity

The deteriorating macroeconomic environment resulting from the crisis is evidenced in the degree of sparsity observed along the prediction period. For NGDP and UNEMP, the number of forecasters included in the combinations began to increase in the second and third quarter of 2008, respectively and their positive trend was maintained until the end of the subsample (see, Figure 2).

For TBILL and HOUSING, the number of forecasters began its growth since the second quarter of 2005 and peaked in June 2006. For the first variable the number of forecasters' remains in its peak value until the end of the subsample; while for HOUSING after shown a declining trend until the last quarter of 2007, it reports an increase during 2008. Recall that the behaviour of TBILL is related to the monetary policy and HOUSING reflects the state of the business cycle earlier than others indicators (see, for instance, Leamer, 2009). Finally, the number of forecasters for CPI and BOND remain unchanged during the crisis period showing a slight increase at the end of the sample.

5. Conclusions

As was stated by Clemen (1989) and confirmed by a large number of empirical studies "Forecast accuracy can be substantially improved through the combination of multiple individual forecasts". However the question of how to combine in a better way the forecasts available is still an open issue.

In this paper we compare the performance of different techniques to combine and to combine selected single forecasts from the SPF. The empirical results suggest that combination schemes that perfom a forecaster's selection yield predictive gains over the widely used summarising measure, the simple average of the survey participants. In particular, the selection process implemented by the SPLS method provides a good prediction performance. The final SPLS forecast combinations are sparse; which implies that forecasters with redundant information or small predictive power over the target are removed from the panel, reducing the estimation error and, therefore, improving the forecast efficiency.

Considering the fourth subsamples analyzed (including the extended sample 2) for the whole period 1991-2012, we find that the SPLS model outperforms the alternative methods in almost 75% of the cases. It is important to highlight that the extended sample 2 includes especifically the crisis period 2008-2009. According to the robustness of the results obtained in the different subsamples and the detailed results for the fourth subsample, the model was able to capture the changes in the behaviour of the variables resulting from the financial crisis and performed well during its development, with the exception of HOUSING and BOND. We consider that the prolonged stagnation of housing sector, the severity of the crisis and the sluggish economic recovery have influenced the deterioration the performance of the forecasting models of these latter variables with respect to the average of forecasters.

The economic uncertainty caused by the financial crisis can be evidenced in the change that occurs in the number of panelists who are selected by the model over the prediction period 2005-2009. For NGDP and UNEMP the number of the panelists selected to build the final combination increased during the recession period 2008-2009. For HOUSING a similar behavior was observed temporarily in accordance with the sector cycle and was repeated, with less intensity, during the recession of the economy.

Figure 1
Squared Forecast Errors

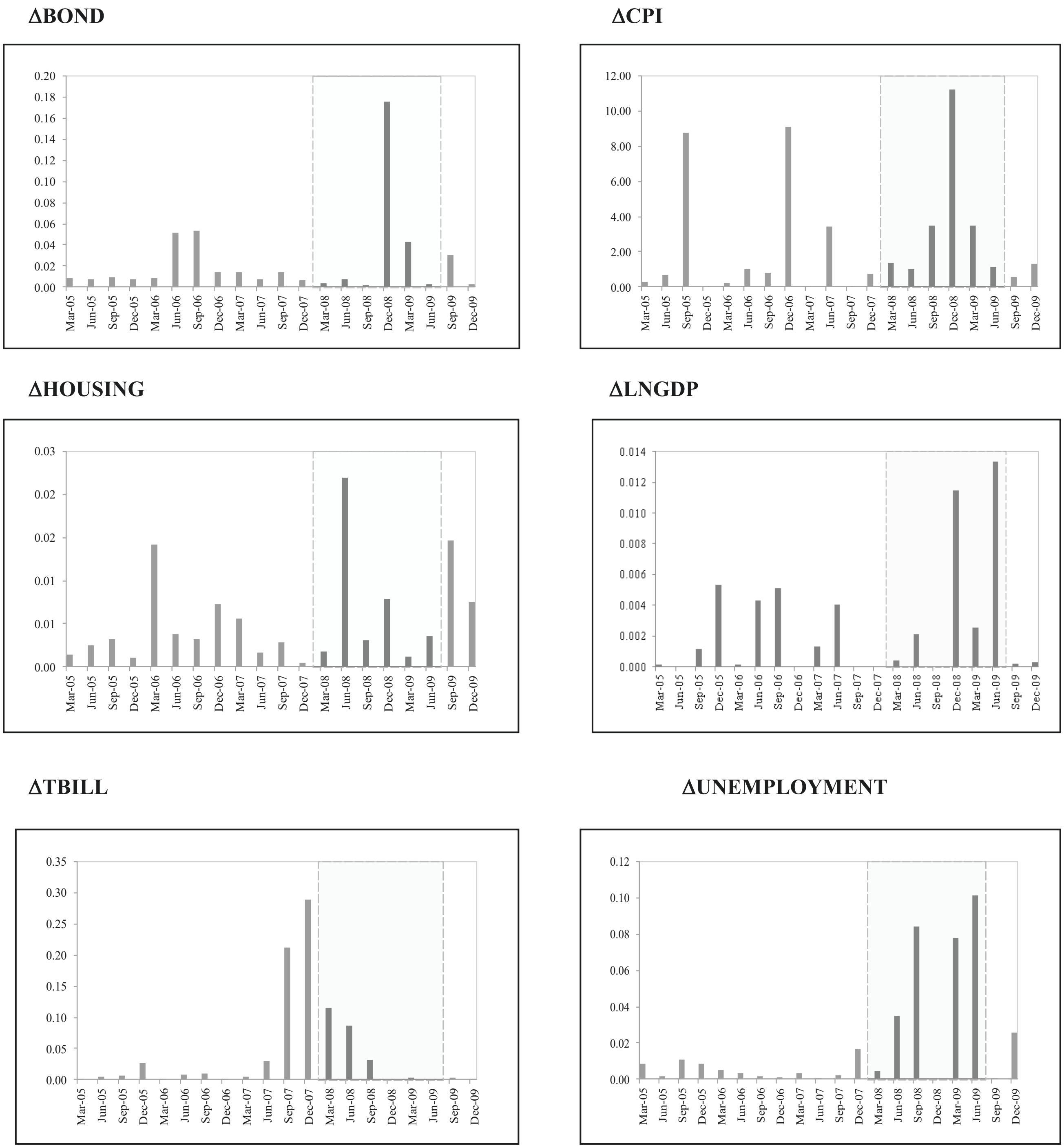
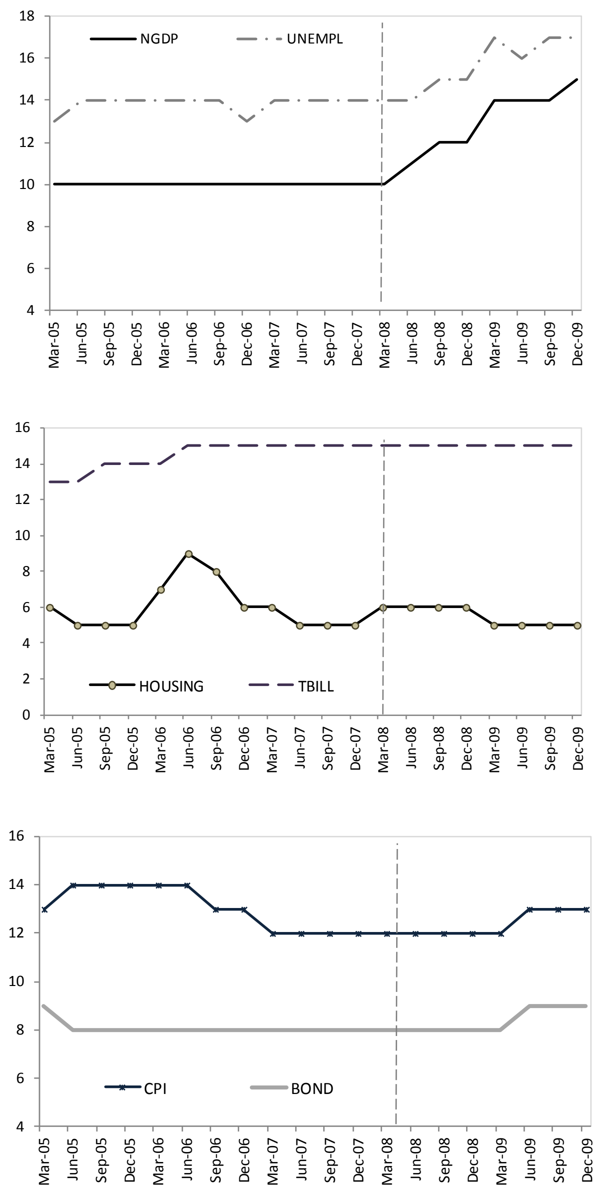


Figure 2
Number of Forecasters Included in the SPLS Combination



References

Aiolfi, M., Capistrán, C. and A. Timmermann (2011). Forecast Combinations. In *Oxford Handbook of Economic Forecasting,* Clements M. and Hendry, D. (eds). Oxford University Press; 355-388.

Aiolfi, M. and C. Favero (2005). " Model Uncertainty, Thick Modelling and the Predictability of Stock Returns". *Journal of Forecasting*. Vol 24 (4): 233–254.

Bates, J. and C. Granger (1969). "The Combination of Forecasts". *Operational Research Quarterly*, 20 (4): 451-468.

Clemen, R. (1989). "Combining Forecasts: A Review and Annotated Bibliography". *International Journal of Forecasting*, 5 (4): 559-583.

Clements, M and D. Harvey (2009). Forecast Combination and Encompassing. In *Palgrave Handbook of Econometrics,* Applied Econometrics, Mills T. and Patterson, K. (eds). Oxford Palgrave, (2):169-198.

Chun H. and S. Keles (2010). "Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection", *Journal of the Royal Statistical Society.* Series B, Statistical Methodology, 72(1): 3-25.

Conflitti, C., De Mol, C. and D. Giannone (2012). "Optimal Combination of Survey Forecasts". ECARES Working Paper 2012-023.

Diebold, F. and P. Pauly (1990). "Structural Change and the Combination of Forecasts". *International Journal of Forecasting,* 6 (4): 503-508.

Fuentes, J., Poncela, P. and J. Rodríguez (2014). “Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting”. *Journal of Applied Econometric,* Forthcoming.

Genre, V., Kenny, G., Meyler, A. and A. Timmermann (2013). “Combining Expert Forecast: Can Anything Beat the Simple Average?”. *International Journal of Forecasting*, 29:108-121.

Granger, C. and Y. Jeon (2004): “Thick modeling”. *Economic Modelling*, 21: 323-343.

Groen, J. and G. Kapetanios (2008). “Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting”, Federal Reserve Bank of New York Staff Report 327.

Hastie T., Tibshirani, R. and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics, 2nd. edition.

Kelly, B and S. Pruitt (2012). “The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors”, Chicago Booth Paper No. 11-19. Fama-Miller Center for Research in Finance.

Leamer, E. (2009). “*Macroeconomic Patterns and Stories. A Guide for MBAs*” Springer-Verlag Berlin.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and R. Winkler (1982). “The Accuracy of Extrapolation (time series) Methods: Results of a Forecasting Competition”. *Journal of Forecasting*, 1: 111-153.

Marcellino M. (2004). “Forecast Pooling for European Macroeconomic Variables”. *Oxford Bulletin of Economics and Statistics*, 66(1): 91-112.

Mincer, J. and Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations:* Analysis of Forecasting Behavior and Performance. Mincer, J. (ed). New York: National Bureau of Economic Research: 1-46.

Poncela, P., Rodríguez, J., Sánchez-Mangas R. and E. Senra (2011). "Forecast Combination through Dimension Reduction Techniques". *International Journal of Forecasting,* 27: 224-237.

Samuel, J. and R. Sekkel (2013). "Forecasting with Many Models: Model Confidence Sets and Forecast Combination". Working Paper 2013-11, Bank of Canada.

Stock J. and M. Watson (2001). "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series". In *Cointegration, Causality, and Forecasting: Festschrift in Honour of Clive W.J. Granger,* R.F. Engle and H. White (eds). Oxford University Press: 1-44.

Stock J. and M.Watson (2004). "Combination Forecasts of Output Growth in a Seven-Country Data Set". *Journal of Forecasting*, 23: 405-430.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society*. Series B (Methodological), 58(1): 267-288.

Timmermann, A. (2006). "Forecast Combinations". In *Handbook of Economic Forecasting*, Elliot, G., Granger, C. and A. Timmermann (eds.). North-Holland Vol I; 136-196.

Wold, H. (1966). "Estimation of Principal Components and Related Models by Iterative Least Squares". In *Multivariate Analysis,* Krishnaiaah PR. (ed.), Academic Press: New York; 391-420.