# Nonlinear Forecast Combinations: An Example Using Euro-Area Real GDP Growth*

Heather D. Gibson
Bank of Greece

Stephen G. Hall
University of Leicester
Bank of Greece

George S. Tavlas†
Bank of Greece
University of Leicester

July 2018

ABSTRACT

The forecasting literature shows that when a number of different forecasters produce forecasts of the same variable it is almost always possible to produce a better forecast by linearly combining the individual forecasts. Moreover, it is often argued that a simple average of the forecasts will outperform more complex combination methods. This paper shows that, analytically, nonlinear combinations of forecasts are superior to linear combinations. Empirical results, based on comparisons of real GDP growth projections with outturns for the euro area using time-varying-coefficient estimation, confirm that analytical result, especially for periods marked by structural changes.

Keywords: Nonlinear Forecast Combinations, Nonlinear Models, Time-varying Coefficients

JEL classification: E37, E53, C45

* We have benefited from very helpful comments from Fredj Jawadi and two referees.
† Corresponding author: George Tavlas, Bank of Greece, 21 E Venizelos Ave, Athens, 10250, Greece, Tel. no. +30 210 320 2370; Fax. no. +30 210 320 2432; Email address: gtavlas@bankofgreece.gr.

*1. Introduction*

It is often the case that policy makers have a number of different forecasts available to them for any particular variable of interest. In such situations, it would seem to be inefficient to use information provided by a single forecast while ignoring the information contained in the other forecasts. Consequently, a technique that allows these different forecasts to be used in such a way that produces a superior forecast would have substantial value from the policy makers' perspective. Beginning with the work of Granger (1969), who argued that forecast accuracy can be improved by using a covariance method of combining forecasts rather than any individual forecast, and the work of Granger and Ramanathan (1984), who proposed a regression-based methodology for combining forecasts, a large literature has established the overall superiority of a linear combination of forecasts. The intuition underlying this finding is straightforward: combining forecasts achieves diversification gains -- combinations of forecasts based on, say, different information sets, pool together different sources of information and, therefore, should result in lower expected loss (for example, lower expected bias). Moreover, given that it is likely that some models will be misspecified over certain periods of time -- for example, some models may adjust faster to regime changes than other models -- combining forecasts from a number of models may offer some insurance against breaks or other unknown sources of misspecification (Elliott and Timmermann, 2008, p. 42; González-Rivera, 2013, Chapter 9). The empirical literature on forecast combinations has, however, focused exclusively on linear combinations. Yet, in real-world situations we would expect that linear forecast combinations are often misspecified; since a linear combination is always a special case of a nonlinear combination, the nonlinear combination must perform at least as well as the linear one.

In this paper we aim to extend the literature on the combination of forecasts by applying forecast combinations to a nonlinear context. We show that combinations of forecasts based on weights derived using time-varying-coefficient (TVC) and neural nets, both of which are particularly adept at dealing with periods of structural change, produce superior forecasts compared to combinations derived using a linear framework. Specifically, we combine forecasts of real GDP growth for the euro area based on TVC and neural nets and compare those combinations with combinations generated with linearly-derived weights. The results indicate that forecast accuracy is generally improved using the nonlinear combinations.

The remainder of the paper consists of five sections. Section 2 presents a framework demonstrating that nonlinear combinations of forecasts are analytically superior to linear combinations. Section 3 describes our data and the models used to forecast real GDP. Section 4 provides the empirical results. Section 5 provides a simple intuitive experiment that illustrates the differences in results generated by linear and nonlinear forecasting techniques. Section 6 presents concluding remarks. An Appendix explains our treatment of interpolation.

## *2. Nonlinear Forecast Combinations*

As mentioned, a large literature has emerged showing that linear forecast combinations are generally superior to individual forecasts. Stock and Watson (1999) found that pooled forecasts outperformed forecasts for US macroeconomic series compared with any single method, including neural networks and autoregressions. Hibon and Evgeniou (2005), using a database consisting of some three thousand time series, found that it is “less risky” (in terms of the variance of the forecasts around the actual outcomes) to combine forecasts than to use an individual forecasting method. Genre *et al.* (2013) analyzed the gains from combining the forecasts from the ECB Survey of Professional Forecasters. Those authors found that, for real GDP growth and the inflation rate, only a few of the individual forecast combination schemes considered (*e.g.,* least squares estimates of optimal weights, Bayesian shrinkage, principal components and trimmed means) outperformed the simple equally-weighted average forecast. Franses, McAleer, and Legerstee (2012) and Elliott and Timmermann (2016, Chapter 14) provided surveys of the literature on forecast combinations.

It is straightforward to show why combining forecasts will generally provide a more accurate forecast than any of the individual forecasters in the combination. . Consider two forecasters, a and b, both of whom make a forecast of a variable $\hat{y}$ at time t for a future period t+h. We can combine these two forecasts with weights that sum to unity to give forecast $\hat{y}^c$.

$$\hat{y}^c{}_{t+h,t} = \alpha \hat{y}^a{}_{t+h,t} + (1-\alpha)\hat{y}^b{}_{t+h,t} \tag{1}$$

where $\hat{y}^a_{t+h,t}$ is the forecast made at time *t* by forecaster *a* for period *t+h* and $\hat{y}^b_{t+h,t}$ is the analogous forecast made by forecaster *b*. The error made by each individual forecaster is $e^a \, and \, e^b$ so the error made by the combined forecast will be:

$$e^c{}_{t+h,t} = \alpha e^a{}_{t+h,t} + (1-\alpha)e^b{}_{t+h,t} \tag{2}$$

where $e^{c}_{t+h,t}$is the combined forecast error made at time $t$ for period $t+h$ and the variance of the combined error will be:

$$\sigma^2{}_c = \alpha^2\sigma^2{}_a + (1-\alpha)^2\sigma^2{}_b + 2\alpha(1-\alpha)\sigma_{ab} \tag{3}$$

where $\sigma_\alpha^2$ and $\sigma_b^2$ are the variances of the forecast errors of forecasters a and b, respectively, and $\sigma_{ab}$ is the covariance. Unless the correlation between the two forecast errors is unity the combined forecast is likely to have a smaller variance than either of the individual forecasts. Bates and Granger (1969) showed how optimal weights could be calculated from the relative variances and covariances, while Granger and Ramanathan (1984) proposed a simple regression combination method, which turns out to give the same, optimal weights as under the Bates and Granger procedure. In the simple regression combination approach, we estimate the following OLS regression:

$$y_{t+h,t} = \beta_0 + \beta_1\hat{y}^a{}_{t+h,t} + \beta_2\hat{y}^b{}_{t+h,t} + \varepsilon_{t+h} \tag{4}$$

The estimated coefficients give the optimal weights. The constant in the regression allows for the possibility that either of the two forecasts may be biased; the two coefficients are no longer forced to add to unity (although this constraint could be imposed).

The point made above, that equal weights often perform better than optimal ones, is often explained on the basis of structural breaks in the weights. This argument implies that allowance should be made for these breaks using some form of nonlinear model.  the existing literature, however, has, by-and-large, focused on forecast combinations within a linear framework. In what follows, we allow for nonlinear combinations of forecasts using two empirical approaches -- (1) time-varying-coefficients (TVCs) and (2) the neural net; we describe these approaches in the next section. Our conjecture is that TVC-based nonlinear combinations will produce forecasts that are at least as good as linear combinations, especially if the variables under consideration have been subjected to structural change. A related conjecture is that the neural nonlinear combinations would perform as well or better than linear combinations if the variables under consideration follow a stable nonlinear structure.

### *3. Data and Empirical Methodology*

To examine the foregoing conjectures, we use quarterly forecasts of GDP growth for the euro area. Our data set covers the quarterly period 2000:Q1 through 2016:Q3. However, the estimation sample period is 2002:Q1 to 2016:Q3 to account for initialization in several of the techniques that we will employ.[1] The data set we use is taken from the European Central Bank's (ECB's) quarterly report, *Survey of Professional Forecasters.*[2] This survey reports forecasts made by approximately 40 individual private forecasters. The *Survey* does not, however, contain a complete set of predictions for each of the forecasters during the entire estimation period. There are gaps reflecting the fact that individual forecasters did not consistently provide responses, while, over time, some forecasters moved in, while others moved out, of the *Survey.* We require a reasonably complete set of forecasts made by individual forecasters over our estimation period. We have, therefore, chosen the largest possible subset of the full group of forecasters which gives

[1] To make the estimation periods comparable, we also dropped observations from the non-recursive procedures discussed in what follows.

[2] The *Survey* presently includes predictions made by some 40 forecasters. The number of forecasters has not been constant; it has been rising over time.

us a set of nearly complete forecasts. This procedure yielded predictions made by six forecasters. Even the forecasts provided by these forecasters did not yield an entirely complete set of forecasts as there were individual quarters missing for some forecasters.[3] In order to deal with these missing observations we interpolated the data set using a cubic spline interpolation method. The Appendix shows that the actual method of interpolation makes very little difference.

We now describe our empirical methodology in two steps. First, we provide a description of the models used. Second, we then outline our recommended modelling strategy.

### *3.1 The Models*

The following estimation methods are used to generate forecast combinations.

*(1) Simple averages*

We take the average of the forecasts of each of the forecasters. At each point of time we average the forecasts at that particular point of time to produce a combined forecast. Thus

$$\hat{y}^{av}_{t+h,t} = \frac{1}{6}\sum_{i=1}^{6}\hat{y}^{i}_{t+h,t} \tag{5}$$

$\hat{y}^{i}_{t+h,t}$ is the forecast of the GDP growth rate in period *t+h* made in period t by forecaster i, i=1…6 and $\hat{y}^{av}_{th,t}$ is the combined forecast given by averaging the individual 6 forecasters together.

*(2) OLS*

[3] However, the missing quarters were relatively few.

We run an OLS regression of the actual GDP growth rate (the dependent variable) on the six forecasts for the entire sample period.

$$y_{t+h} = \beta_{0,ols} + \sum_{i=1}^{6} \beta_{i,ols} \hat{y}^{i}_{t+h,t} + e_t \qquad (6)$$

where $y_{t+h}$ is the actual growth rate of GDP in period *t+h*. This procedure provides a set of OLS weights $\beta_{i,ols}\, i = 0...6$ . For each quarter, we then generate a new combined forecast using the OLS weights.

$$\hat{y}^{ols}_{t+h,t} = \beta_{0,ols} + \sum_{i=1}^{6} \beta_{i,ols} \hat{y}^{i}_{t+h,t} \qquad (7)$$

This combined forecast becomes our new forecast based on the whole sample OLS weights. This forecast is not, of course, available in real time since we cannot know these weights until we have the full sample of data. This procedure does, however, serve a useful role as a benchmark since these weights will, by construction, yield the best fitting (in a minimum mean square error sense) combination of forecasts that is possible within a linear constant parameter framework.

*(3) Recursive Least Squares (RLS)*

The OLS weights are derived on the basis of the entire sample. Thus, at each point of time they incorporate future values to derive weights for the present period. To deal with this situation, in which future information is not available in real time, we use RLS. Under RLS, we begin with an initial sample of data and re-estimate the coefficients period-by-period as we add additional observations until the sample is exhausted. We construct our RLS combined forecast as follows.

$$\hat{y}^{rls}_{t+h,t} = \beta_{0,rls} + \sum_{i=1}^{6} \beta_{i,rls} \hat{y}^{i}_{t+h,t} \tag{8}$$

*Nonlinear combinations.* We now outline the two classes of nonlinear models we will use -- that is, the time varying coefficient (TVC) model and a neural net. There are, of course, many forms of non-linear models that we might use. Our choice of these two particular models rests on an important distinction between (i) nonlinearity and (ii) structural change (or a break in parameters). A break in a set of parameters has to be defined with respect to a particular parametric model. Thus, if we have a linear model we might find the parameters break at a particular point but there may be a nonlinear model where the nonlinearity would allow for the apparent break in the linear model and so the nonlinear model would have constant parameters. The two nonlinear models we have chosen are in a sense at the two extremes of this dichotomy. The TVC model works within a linear framework except the parameters are allowed to break at every period. The neural net is able to mimic any form of constant parameter nonlinear structure up to any desired level of accuracy. In the limit both models are able to capture any process. However, we believe that, if a relatively simple but unknown nonlinear structure which is stable exists, then the neural net will be a more parsimonious model to use for the optimal combination. If, on the other hand, a very complex nonlinear structure is required before we are able to achieve constant parameters, then the TVC model may be a better tool.

(4) *Time Varying Coefficient (TVC) Models.*

The TVC approach to dealing with the foregoing problem proceeds from an important theorem that was first established by Swamy and Mehta (1975), and which was subsequently confirmed by Granger (2008). This theorem states that any nonlinear functional form can be exactly represented by a model that is linear in variables, but which has time-varying coefficients. The implication of this result is that, even if we do not know the correct functional form of a relationship, we can nevertheless represent this relationship as a TVC relationship and, thus, estimate it. Hence, any nonlinear relationship may be stated as:

$$y_t = \gamma_{0t} + \gamma_{1t}x_{1t} + \ldots + \gamma_{pt}x_{pt} \tag{9}$$

Thus, our general TVC nonlinear forecast combination equation may be stated as:

$$\hat{y}^{tvc}_{t+h,t} = \gamma_{0t} + \gamma_{1t}\hat{y}^{1}_{t+h,t} + \ldots + \gamma_{6t}\hat{y}^{6}_{t+h,t} \tag{10}$$

and the time-varying-coefficients may be estimated from the following state space form using the Kalman filter formulation:

$$y_{\ t+h,t} = \gamma_{0t} + \gamma_{1t}\hat{y}^{1}_{t+h,t} + \ldots + \gamma_{6t}\hat{y}^{6}_{t+h,t} \tag{11}$$

which is the measurement equation for the state space form. There are several possibilities for the structure of the state equations. The simplest is:

$$\gamma_{it} = \gamma_{it-1} + \varepsilon_{it} \quad i = 0,\ldots,6 \tag{12}$$

That is, the parameters are specified as a simple random walk. In the section presenting our results, we will call this model TVC1. This specification allows the parameters to change in an unrestricted way; however, the forecast for the future values of the parameters would always be a constant value at the last period's estimate, with that estimate changing as the sample increases.

If the parameters had been changing in a steady or trend-like way, we would want to incorporate this trend behavior in our forecast. This can be done in the following way:

$$\begin{aligned} &\gamma_{it} = \gamma_{it-1} + \gamma_{6+1+it-1} + \varepsilon_{it} \\ &\gamma_{6+1+it} = \gamma_{6+1+it-1} \qquad i = 0,\ldots,6 \end{aligned} \tag{13}$$

This procedure puts a local stochastic drift term into each of the parameters, allowing each parameter to continue changing into the future. In the results section we will call this variant TVC2.

In using the Kalman filter models, values have to be assigned to the variances of both the measurement and the state equations. This is typically done by maximising the likelihood function with respect to these parameters. However, if this is done over the full sample, the procedure would be subject to the criticism that the estimated parameters would not be available in real time (much like the criticism applicable to using full-sample OLS). There are two valid approaches for addressing this issue. First, we could estimate the variances using recursive maximum likelihood, but this approach suffers from the problem that we would need a reasonably large start-up sample for the first set of estimates. Second we can impose these parameters 'a priori'. We have chosen this latter approach, and we impose the variance of the measurement equation at zero and the variance of each of the state equations at unity. The rationale for our approach lies in the Swamy and Mehta theorem which states that the TVC equation has no errors; hence, the variance is zero and the smoothed TVC parameters provide a perfect fit. The filtered and forecasted parameters will, of course, not fit perfectly. In implementing our approach, we also compared the full sample maximum likelihood parameters with the parameters which come from the imposed set of parameters; we found that there is very little difference between the methods.

*(5) The Neural Net*

A neural net is essentially a complex combination of simple nonlinear functions. As such, a sufficiently-complex net can approximate any unknown nonlinear function with any desired degree of accuracy (White, 1992; Swanson and White, 1997).[4] The simple feedforward neural net can be represented as follows:

$$\hat{y}^{nn}_{t+h,t} = h(\sum_{k=1}^{K} \alpha^{nn}{}_{k}\, g(\sum_{j=1}^{6} \beta^{nn}{}_{jk} \hat{y}^{j}_{t+h,t})) \qquad (14)$$

where $\hat{y}^{nn}_{t+h,t}$ is the combined forecast for GDP growth for *t+h* formed in period *t* by the neural net method, $\hat{y}^{j}_{t+h,t}$ is the forecast for GDP growth made by forecaster *j* in period *t* for period *t+h*, $\alpha^{nn}$ *and* $\beta^{nn}$ are parameters, and *h* and *g* are simple nonlinear functions; in practice *h* and *g* are usually the same function. In our application, both *h* and *g* are the logistic function, $g(u) = h(u) = 1/(1+e^{-u})$. This specification is called a feed-forward model since $\hat{y}^{j}_{t+h,t}$ affects $\hat{y}^{c}_{t+h,t}$ but $\hat{y}^{c}_{t+h,t}$ does not affect $\hat{y}^{j}_{t+h,t}$. This specification is a simple one-layer neural net with J inputs and K neurons.[5]

From an econometric perspective, as the net becomes more complicated the parameters, $\alpha^{nn}$ *and* $\beta^{nn}$, grow in rapidly number, and are generally not identified or unique. The process of arriving at estimates of these parameters is termed learning (or training) and it is generally undertaken by some form of hill-climbing technique that attempts to minimize

[4] A good introduction to the neural net method is found in Kuan and White (1994).
[5]A more complex model would involve further summations over more nonlinear transformations.

the squared deviations between actual GDP growth and the combined forecast from the net. Typical procedures are back propagation, simulated annealing or genetic algorithms. It is generally recognized that a complex net will need a large number of observations to train it effectively. Given our limited number of observations, we restrict the net to one hidden layer and six neurons (our six forecasters).

In most of the neural-network literature, applications divide the sample into a training sub-sample and a forecasting sub-sample. This procedure does not represent how a neural net would be used in actual forecasting, since in a practical setting we would use all available data to train the net and then forecast into the future. For this reason, we will implement a recursive neural net, training it initially over an initial small sample and then recursively moving through the sample forecasting and re-training in a manner analogous to recursive OLS.

It is important to point out that the net is based on the assumption of the existence of a stable nonlinear structure. A sufficiently complex net can approximate a changing structure but this specification may have to be excessively complex. Essentially, the point is that $\alpha^{nn}$ *and* $\beta^{nn}$ are being trained over the whole sample up to period t-1. Therefore, information from the past is not being written off[6] as it is with the TVC model above.

*3.2. Discussion*

The relative performance of the above models will depend on the type and form of structural instability that may be present in the real world. If we begin by defining

[6] That is, the parameters of the neural net are constant up to period t; thus, if there is a break over this period, the parameters are determined with equal weights over the period.

structural instability with respect to the standard fixed parameter linear model, then, if this model is stable, RLS is the best forecast combination weights that can be used in a real time forecasting exercise; therefore, the full sample OLS weights are the best weights that could be used. If, however, the standard linear model is subject to unstable parameters, and if the real world can be characterised by a relatively-simple, but unknown, nonlinear relationship with stable parameters, we would expect the neural net to perform well. However, if structural instability is endemic such that it needs an extremely complex nonlinear model to capture it with stable parameters, then we would expect the TVC model to be the better approach. Our broad strategy is to begin by assessing the stability of the standard linear model and then go on to see if the above expectations as to the best model are born out.

We use the conventional CUSUM squared test to assess structural instability. We focus on the CUSUM squared test because our aim is to learn about structural breaks rather than to conduct a classical hypothesis test. If there is no sign of instability, we would expect the conventional combination techniques to work well. However, if there are signs of instability in the weights we would expect the TVC based methods to outperform the fixed-weight methods, that is averaging or OLS based combinations. If the recursive estimates of the weights seem to be trending, we would expect TVC2 to outperform TVC1; if not, then TVC1 would likely be the better performer. If there is little indication of instability then we would still expect the neural net to outperform the linear combination methods, although, as noted above, the neural net may not perform well during episodes of extreme structural instability.

## *4. Empirical Results*

Figure 1 shows actual euro-area real growth, quarter-over-previous quarter (annualized) over the period 2001:Q1 through 2016:Q3. Overall, real growth averaged 1.3 per cent during the period. However, there was considerable variation during sub-periods; growth averaged 2.3 per cent during 2001:Q1 to 2007:Q4; with the onset of the international financial crisis (which, initially, had little effect on the euro area) and the outbreak of the euro-area sovereign debt crisis in late-2009, real growth swung into negative territory, at minus 0.1 per cent, on average, during 2008:Q1 to 2013:Q4. Then, abetted by the ECB's non-standard monetary measures, including quantitative easing, growth returned to positive territory, averaging 1.8 per cent during 2014:Q1 to 2016:Q3.

To evaluate forecast accuracy, we use three criteria: (i) root mean square error, which is the square root of the mean of the squared differences between predicted and actual values; (ii) mean absolute error (MAE), which is the mean of the absolute errors (the RMSE and MAE are different because the squaring of the differences used to calculate the RMSE gives larger weights to large errors in the mean calculation); and (iii) mean absolute percent error (MAPE), which is the mean of the absolute per cent errors -- taking per cent errors weights the errors differently for different inflation rates (for example, a 1 percentage point error is a large per cent error when actual inflation is zero, and small when inflation is high.)

Table 1 reports results for euro-area GDP growth based on simple averages of forecasts and forecasts based on OLS weights. The table also presents the RMSEs, the MAEs, and the MAPEs for the six individual euro-area forecasters used in this study. As shown in the table, each of the individual forecasters performs similarly. For example, individual-forecaster RMSEs fall within the narrow range of 1.882 (forecaster 4) to 1.928 (forecaster

2). The equally-weighted combined forecast shows little improvement and is, in fact, less accurate than the forecasts of three of the individual forecasters. The OLS combination does, however, represent an improvement on the basis of the RMSE and MAE; for example, the RMSE falls to 1.861. As noted, however, the OLS weights are derived on the basis of the entire sample period and so, at each point in time, they use (future) actual inflation to derive the weights for the current period. This procedure would obviously not be practical in an actual forecasting situation. Thus, the increase in accuracy for the OLS combination is likely to be misleading.

To deal with the problem arising from the use of future outcomes, we turn to recursively-estimated weights. To provide a picture of the evolution of these weights, Figure 2 displays the recursive coefficients for each of the six forecasters used in our sample. As shown in the figure, the recursively-estimated coefficients are subject to considerable instability. Figure 3 shows the CUSUM of squares test for parameter instability; the CUSUM test is known to be a fairly weak test, but even this weak test shows that the parameters significantly fail the stability test during the period around the outbreak of the international financial crisis in 2008 and during the height of the euro-area sovereign-debt crisis in 2010-2012.

Table 2 shows a comparison of the forecasts using equal weights, full sample OLS and recursive parameters, with the former two forecast results carried over from Table 1.

In this case the recursive parameters perform much worse than the full sample OLS parameters, under the RMSE and MAE criteria; for each of our measures of accuracy the equal weights actually outperform the recursive OLS weights. Clearly, the unstable nature of the OLS weights causes a deterioration of forecasting performance. This outcome motivates us to turn to the time-varying-coefficient technique, which will allow us to account for coefficient instability.

The results using the two TVC1 models and the nonlinear neural net model are presented in Table 3. For convenience of comparison, the table also carries over the results from Table 2 based on the methods of equal weights and recursive OLS. The TVC results represent a considerable improvement -- of the order of 30 per cent (based on RMSEs) over both the recursive OLS results and the equally-weighted results. Thus, there appears to be considerable forecasting gains in using time-varying weights.

TVC2 does not perform as well as the simpler TVC1 method; nevertheless, it still outperforms recursive OLS. The intuitive reason underlying the superior performance of TVC1 relative to TVC2 is that the combination weights are relatively stable and mean-reverting so that projecting them as a level process (*i.e.,* a pure random walk), as under TVC1, produces a better result than projecting their recent trend movement into the future (*i.e.,* as a random walk with drift), as under TVC2. The neural net yields essentially no improvement over the methods based on recursive least squares and equal weights. This finding is attributable to the high volatility of the actual growth rate of GDP and, as we are forecasting four periods into the future, the net seems to capture current behavior but is not able to capture future behavior. Moreover, although the net is a nonlinear procedure, it is also a fixed-parameter procedure over the estimation sample. Consequently, it seems that the underlying assumption of structural stability is not valid, leading to essentially no improvement in forecasting accuracy compared with the linear combination methods.

*5. Some Intuition: A Simple Experiment*

To illustrate the important difference between OLS, recursive OLS, and the TVC techniques in the presence of structural breaks, we have set up a simple experiment. We create an artificial series which takes the value 1 for the first half of the sample -- data points 1 through 8 -- and then is subjected to a break in that it changes value to a value of 2 for the second half of the sample -- data points 9 through 16. This line is the solid line labelled "data" in Figure 4. We then estimate a simple OLS regression of the form:

$$Data = c$$

where $c$ is a conventional constant term. This produces a value of 1.5 for $c$, (the OLS line in Figure 3) which would be our full sample estimate of the OLS coefficient. Clearly, this is an average of the two regimes, and it gets neither right.

We then perform recursive OLS on this regression. The recursive parameters get the first regime right (its line overlays the solid line representing the actual data for data points 1 through 8), but once the break occurs the recursive parameter moves towards the full sample OLS parameter slowly, and never gets the correct answer for the second regime.

We then perform a TVC estimation using the Kalman filter with a coefficient determined by a random walk of the form:

$$Data = c_t + \varepsilon_t$$

where the state equation is

$$c_t = c_{t-1} + \phi_t$$

The Kalman filter works as follows. It first filters the data, producing estimates for the coefficient at time t based on data from 1 to t-1. Then it smooths the coefficients based on the full sample, so the coefficient for period t is based on data from 1 to T. We report both the filtered (or predicted) and smoothed coefficients in Figure 4 since the smoothed coefficients are equivalent to full-sample OLS, while the filtered coefficients are equivalent to recursive OLS. As shown in Figure 4, *TVC smoothed captures the break perfectly; it gets it right from the instant the break occurs b*ecause it knows the break will happen. This circumstance reflects the fact that this method uses future information. In contrast, *TVC predicted (the filtered estimates)* misses the break at the instant the break occurs. This circumstance reflects the fact that this procedure does not use future information. As shown in Figure 4, however, the parameter under *TVC predicted* moves quickly to its true value in the second regime.

This simple experiment illustrates that, while recursive OLS can be used to test parameter stability, it is not a consistent estimator of a parameter subject to a structural break. In contrast, TVC estimation can give consistent estimates of a parameter in the presence of a structural break.

## *6. Conclusions*

The ability to provide accurate forecasts of real GDP growth plays a crucial role in a central bank's decisions about monetary policy. Forecast errors will lead to inappropriate policy choices, destabilizing output relative to potential output, and inflation relative to target inflation. A consensus has emerged in the literature that the forecast accuracy of macroeconomic time series can be improved, thus reducing the possibility of inappropriate policy choices, through the application of linear combinations of forecasts.

Although linear forecast combination has long been viewed as a way to improve forecast accuracy, if there are serious structural breaks in the OLS weights used to form the combination then recursive estimation of these weights does not provide a consistent estimate of the true underlying optimal weights at each point in time. We argued that a nonlinear forecast combination, that takes account of structural change, should outperform recursive OLS, except in the unusual case where the recursive parameters are completely stable. We also pointed out that we can represent this unknown optimal nonlinear function by a linear relationship with time-varying parameters. We argued that by using these predicted time varying weights we should be able to produce better combinations of forecasts. We applied this idea to a set of forecasts of euro-area real GDP growth and found that the TVC model combinations perform better -- on the order of about 30 per cent (based on RMSE) -- than linearly-based forecasting methods.

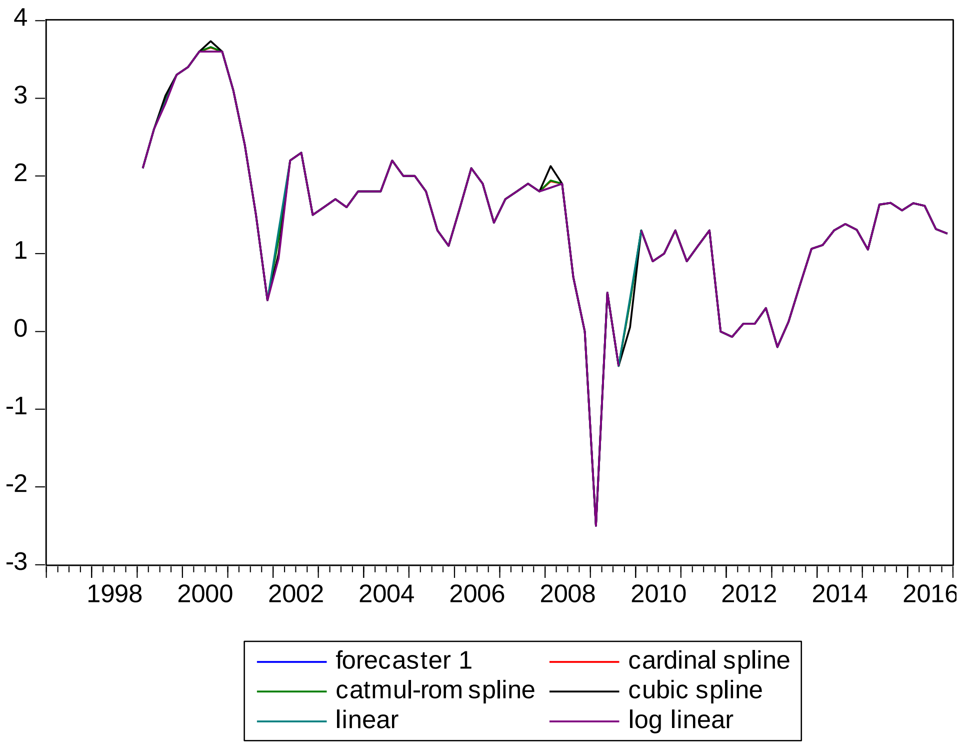
We also applied this idea to a set of forecasts based on the neural net method, an approach that is nonlinear but does not account for structural change. The TVC model combinations also outperformed the neural-net based forecasts -- of the order of 30 per cent (based on RMSEs) -- suggesting that forecast combinations based on TVC are good at dealing with periods of structural change.

We stress that these findings are an initial attempt to deal with the issue of nonlinear forecast combinations. Further work, pertaining to different data samples and alternative nonlinear methods, can shed additional light on the investigation undertaken in this paper.

Appendix: The choice of interpolation method

As discussed in the text, we have interpolated the missing values in each forecast using a cubic spline. The choice of this method, however, has virtually no impact on our final results. We illustrate this by comparing a range of interpolation procedures for the first forecaster in our sample. We employ five commonly-used interpolation methods; linear interpolation, log linear interpolation, a cardinal spline, a cubic spline and the catmul-rom spline. Figure A1 shows the data which emerges from each of these methods, along with the actual. It is clear from this figure that all the methods give very similar values -- thus, the lines are essentially indistinguishable. The difference between them is negligible with respect to the variation in the forecast itself.

Figure A1: Various Interpolation Methods.



In Table A1 we show the actual values for one of the interpolated observations. It is clear that all the observations are very similar.

Table A1: The interpolation methods for the quarter 1999q3

| | forecaster 1 | cardinal spline | catmul-rom | cubic spline | linear | log linear |
|---|---|---|---|---|---|---|
| 1999Q1 | 2.100000 | 2.100000 | 2.100000 | 2.100000 | 2.100000 | 2.100000 |
| 1999Q2 | 2.600000 | 2.600000 | 2.600000 | 2.600000 | 2.600000 | 2.600000 |
| 1999Q3 | NA | 2.980000 | 2.983333 | 3.032536 | 2.950000 | 2.929164 |
| 1999Q4 | 3.300000 | 3.300000 | 3.300000 | 3.300000 | 3.300000 | 3.300000 |
| 2000Q1 | 3.400000 | 3.400000 | 3.400000 | 3.400000 | 3.400000 | 3.400000 |

It would be possible to deal with the missing observations using the Kalman filter itself in the estimation of the TVC model. This would involve setting the variance of the measurement equation to infinity in each observation where there was a missing observation. In practice this would not be a good way forward for two reasons; first it would not solve the problem for the other techniques, recursive OLS equal weights and the neural net. Second, as each forecaster has on average six missing values and there are six forecasters, this would have effectively excluded 36 observations, which would be a very large part of our available sample.

*References*


Bates John, Granger, Clive, 1969. The Combination of Forecasts. *Operations Research Quarterly* 20(4), 451-468.

Elliott, Graham, Timmermann, Allan, 2008. *Economic Forecasting*. *Journal of Economic Literature* 46(1), 3-56.

Elliott, Graham, Timmermann, Allan, 2016. *Economic Forecasting*. Princeton University Press: Princeton, N.J.

Franses,Philip Hans, McAleer, Michael*,* Legerstee Rianne, 2012. Evaluating Macroeconomic Forecasts: A Concise Review of Some Recent Developments. Journal of Economic Surveys, 28(2), 195-208

Genre, Veronique, Kenny, Geoff, Meyler, Aidan, Timmermann, 2013. Combining Expert Forecasts: Can Anything Beat the Simple Average? *International Journal of Forecasting* 29(1), 108-121.

González-Rivera, Gloria, 2013. *Forecasting for Economics and Business*. The Pearson Series in Economics, Pearson/Addison-Wesley.

Granger, Clive, 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37(3), 424-38.

Granger, Clive, 2008. Non-linear Models: Where Do We Go Next – Time Varying Parameter Models. *Studies in Nonlinear Dynamics and Econometrics* 12, 1-10.

Granger, Clive, Ramanathan, Ramu, 1984. Improved Methods of Combining Forecasts. *Journal of Forecasting* 3(2), 197-04.

Hibon, Michele, Evgeniou, Theodoros, 2005. To Combine or Not to Combine: Selecting among Forecasts and Their Combinations. *International Journal of Forecasting* 21, 15-24.

Kuan, Chung-Ming, White, Halbert, 1994. Artificial Neural Networks: An Econometric Perspective. *Econometric Reviews* 13(1), 1-91.

Rossi, Barbara, Sekhposyan, Tatevik, 2014, Forecast Rationality Tests in the presence of Instabilities, with applications to Federal Reserve and Survey Forecasts. Barcelona GSE working Paper no 765.

Stock, James H., Watson, Mark W., 1999. A Comparison of Linear and Nonlinear Univariable Models for Forecasting Macroeconomic Time Series. In *Cointegration, Causality, and Forecasting:* A *Festschrift in Honour of Clive W. J. Granger,* ed. R. F. Engle and H. White. Oxford and New York: Oxford University Press, 1-44.

Swamy, P. A. V. B., Mehta, Jatinder, 1975. Bayesian and non-Bayesian Analysis of Switching Regressions and a Random Coefficient Regression Model. *Journal of the American Statistical Association* 70(351), 593–02.

Swanson, Norman, R. and White, Halbert, 1997. A Model Selection Approach To Real-Time Macroeconomic Forecasting Using Linear Models And Artificial Neural Networks. *The Review of Economics and Statistics*, MIT Press, 79(4), 540-550, November.Timmermann, Allan, 2006. Forecast Combinations. *Handbook of Economic Forecasting* 1, 135–196.White, Halbert, 1992. *Artificial Neural Networks: Approximation and Learning.* Cambridge, Blackwell.

Figure 1: GDP growth in the euro area 2001-2016 (Annualized changes over previous quarter)


Source: European Central Bank, Data Warehouse

Figure 2: Recursive OLS weights: six individual forecasters

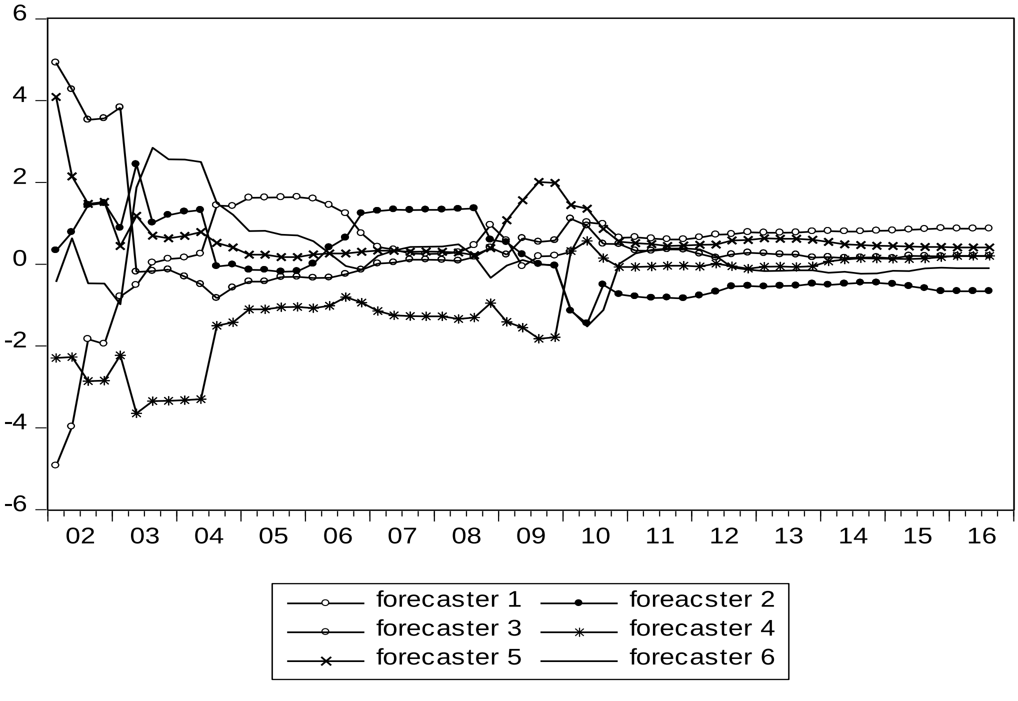


Figure 3: CUSUM of Squares stability test of the recursive coefficients

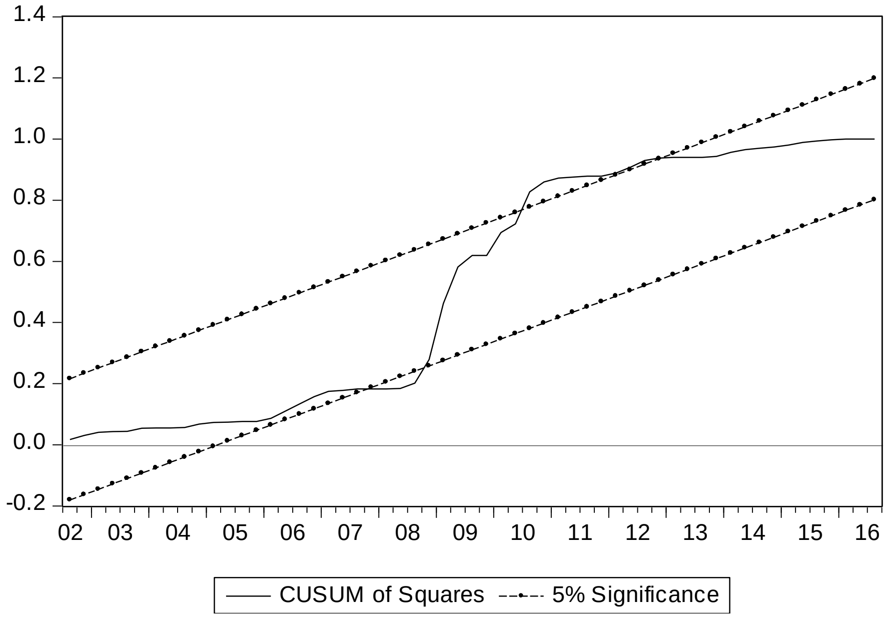



Figure 4: An illustration of the properties of OLS, recursive OLS and TVC

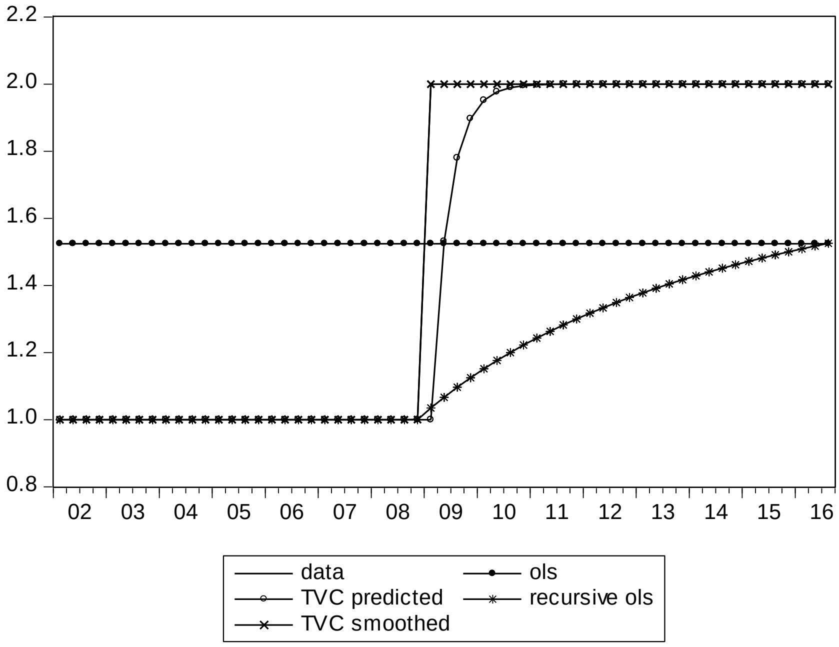


Table 1: Forecasts based on simple averages and OLS weights

| Individual Forecasters | | | | | | | Combined methods | |
|---|---|---|---|---|---|---|---|---|
| | Forecaster 1 | Forecaster 2 | Forecaster 3 | Forecaster 4 | Forecaster 5 | Forecaster 6 | Equal weights | OLS weights |
| RMSE | 1.905 | 1.928 | 1.922 | 1.882 | 1.892 | 1.921 | 1.911 | 1.861 |
| MAE | 1.472 | 1.522 | 1.523 | 1.454 | 1.482 | 1.500 | 1.488 | 1.434 |
| MAPE | 92.26 | 95.65 | 101.97 | 104.6 | 96.6 | 94.6 | 95.48 | 108.8 |

Sample is 2002:Q1 to 2016:Q3

Table 2: Combination Methods: Comparison among equal weights, OLS and recursive OLS

| | Equal weights | Full sample OLS | Recursive OLS |
|---|---|---|---|
| RMSE | 1.911 | 1.861 | 2.138 |
| MAE | 1.488 | 1.434 | 1.839 |
| MAPE | 95.48 | 108.8 | 101.3 |

Sample is 2002:Q1 to 2016:Q3

Table 3: Combination Methods: Comparison among Nonlinear and Linear results

| | Equal Weights | Recursive OLS | TVC1 | TVC2 | Neural Net |
|---|---|---|---|---|---|
| RMSE | 1.911 | 2.138 | 1.507 | 1.503 | 2.185 |
| MAE | 1.488 | 1.859 | 1.164 | 1.222 | 1.836 |
| MAPE | 95.48 | 101.3 | 88.72 | 128.5 | 108.6 |

Sample is 2002:Q1 to 2016:Q3