

Split-then-Combine Method for out-of-sample Combinations of Forecasts

A.S.M. Arroyo¹ & A. de Juan Fernández¹

¹ Avenida Francisco Tomás y Valiente, 5, Mod. E-III-306, 28049 Madrid, Spain

Correspondence: A. de Juan Fernández, Avenida Francisco Tomás y Valiente, 5, Mod. E-III-306, 28049 Madrid, Spain. E-mail: aranzazu.dejuan@uam.es

Received: January 27, 2014

Accepted: February 13, 2014

Online Published: March 5, 2014

doi:10.5430/jbar.v3n1p19

URL: <http://dx.doi.org/10.5430/jbar.v3n1p19>

Acknowledgments: Authors acknowledge funding ECO2012-32854 from DCIGYT. We very sincerely thank the suggestions offered by attendees to the 30th International Symposium on Forecasting, San Diego, CA. Of course, all remaining errors are ours.

Abstract

Relative forecast performance of forecast units may periodically evolve over time. Therefore, it is desirable to take into account their forecast periodicity when forming forecast combinations. When dealing with small samples and small number of models, using panels is an efficient way of pulling out the additional information provided by that periodicity in the data. We capture this periodic information with different weights at different periods that we then keep in the out-of-sample combination. As in the simple average, we do not estimate weights, but instead compute them from panels of forecasts taken as given data. Empirical and bootstrap exercises illustrate the superiority of this method over fixed weight schemes.

JEL Codes: C11, C22, C53

Keywords: Forecast combination puzzle, Period-based weights, Panel decomposition, Changing seasonality, Accuracy measures

1. Introduction

1.1 General

Faced with multiple forecasts of the same variable, an issue that immediately arises is how best to exploit information in the individual forecasts. It is well known, from a simple portfolio diversification argument, that unless one can identify *ex ante* a particular forecasting model that generates smaller forecast errors than its competitors and whose forecast errors cannot be hedged by other models' forecast errors, forecast *combinations* offer diversification gains that make it attractive to combine individual forecasts rather than relying on forecasts from a single model. Its success will depend, however, on how well the combination *weights* can be determined (Timmermann, 2006).

The simplest example of a combination is that of the average of model forecasts. More sophisticated methods of combinations, however, usually do not improve it in empirical applications. Some authors have called this fact as *forecast combination puzzle* (e.g. Stock & Watson, 2004). Many of them consider that this puzzle is due to the instability of the *estimated* weights (e.g. Smith & Wallis, 2009). The usual low accuracy of the combinations based on estimated weights relative to those based on *fixed* weights normally reflects a poor small-sample approximation of optimal population weights by estimated weights. In small samples, estimation of weights requires that the number of individual models, J , be smaller than the number of observations, T_1 , in a training sample. To get close to an optimal weighted forecast, we need a large J , but then, as J increases, we need to estimate an increasing number of weights (*curse of dimensionality*) that works against forecasts based on estimated weights, increasing their MSE when J approaches T_1 . Besides, in the limit when $(J, T_1) \rightarrow \infty$, a feasible bias-corrected *average* forecast is an optimal forecast identical to the conditional expectation (Issler & Lima, 2009).

1.2 Literature Review

While a forecast combination that assumes *constant* weights follows naturally in the case with a time-invariant distribution for the forecasts and realizations, outside this framework it has been useful to consider more general combination schemes by explicitly modeling *time-variation* in the weights: Smooth time-varying parameter model

(Zellner, Hong, & Min, 1991), switching indicator model (Deutsch, Granger, & Terasvirta, 1994) or a mixture model (Elliott & Timmermann, 2005). Stock & Watson (2004) find that the best combination method in terms of MSFE is the simplest one (equal weights) or combinations with weights that change little over time; Yang (2004) developed AFTER (Aggregated Forecast Through Exponential Re-weighting) an algorithm to form time-varying weights in the combination; Zou & Yang (2004), Sánchez (2008), and Wei & Yang (2012) proposed some modifications of the AFTER algorithm; Ravazolo, van Dijk, & Verbeek (2007), and Terui & van Dijk (2002) proposed to estimate the weights using state dependent models and Kalman filter estimation of the weights. Recent proposals are Guidolin & Fangzhou Na (2007) and Guidolin & Timmermann (2009). In these cases, the weights vary according to a switching model with different states, suited to capture the presence of time-variation (structural shifts) in the predictability.

Alternatively, we find that the changing relative forecast performance often encountered in the literature (García-Ferrer, Bujosa, de Juan, & Poncela, 2006) is mainly due to (and we can take advantage of) periodical variations in the data; in particular, it is advisable the use of seasonally unadjusted data (Matas-Mir, Osborn, & Lombardi, 2008). This *periodicity* in the data is in most cases caused by an evolving seasonality that may not be well captured by standard models with stationary seasonality. Whereas seasonal ARIMA models have become standard tools for univariate forecasting, periodically integrated time series models are rarely used in the empirical literature. Apart from examinations of theoretical properties of periodic time series models (Boswijk & Franses, 1996; Franq, Roy, & Saidi, 2011), there are few papers reporting on the performance of these models applied to economic data. In most cases under study, it is shown that periodic models improve the *in-sample* fit considerably whereas they are inferior with respect to *ex-ante* forecasting (Herwartz, 1999). Other times, forecasts improve by using a combination with weights that vary over periods (Novales & Flores de Frutos, 1997).

Our approach may be related to the work on forecast combination (Bates & Granger, 1969; Deutsch *et al.*, 1994; Guidolin & Timmermann, 2009; Granger & Ramanathan, 1984; Hoogerheide, Kleijn, Ravazzolo, van Dijk, & Verbeek, 2010; LeSage & Magura, 1992), dimension reduction (Poncela, Rodríguez, Sánchez-Mangas, & Senra, 2011); periodic ARMA models (Basawa & Lund, 2001; Boswijk & Franses, 1996; Franq *et al.*, 2011; Franses & Paap, 2004; Herwartz, 1999; Jones & Brelsford, 1967; Novales & Flores de Frutos, 1997; Osborn & Smith, 1994), panel-data methods (Issler & Lima, 2009).

1.3 Objective

In this paper, we propose a forecast combination procedure for the case of *periodic* data. Given a set of J forecasts for a (seasonal) time series with M seasons (panels), we split the data into M panels. For each panel, the weights are built on a decomposition of the variability of the forecasts. The *variability* (sum of squares of the difference between each prediction and the overall mean for the panel) is decomposed like in a *two-way fixed effect ANOVA* model, with one factor being the *forecast unit*, and the second factor being the *time*. For each panel, the overall mean, the mean of each forecast unit (along time) and the mean for each time (across models) are calculated. The historical data are treated as fixed. Once the parameters of the ANOVA model (for each panel) are calculated, the weights assigned to each combining forecast are based on the variability explained by the forecast units. The larger the distance between the mean of the predictions of a forecast unit and the overall mean, the larger the weight. The overall mean is the mean of the predictions of all forecast units for that panel. Consequently, actual values of the target variable are not used at all, and the relative prediction performance of the alternative forecast units is not needed at all either since actual values *cancel out* in the definition of the *Split-then-Combine* (STC) weight.

When dealing with *small* samples and *small* number of forecast units, using *panels* is an efficient way of pulling out the additional information provided by that periodicity in the data. As in the simple average, we do *not estimate* the weights (beyond the fact that averages are sample analogues of their population counterparts). Our STC approach uses a panel decomposition of forecasts as an instrument to extract their *fixed* effects. Then, *panel-varying* weights adapt quickly to structural changes or breaks take into account the relative size of each fixed effect to form the STC combination which is kept in the out-of-sample period. In summary, we aim for weights that are (i) Not built upon forecast errors; (ii) Panel-varying; (iii) Not estimated; and (iv) Unchanged in the out-of-sample forecasting.

The paper is organized as follows: In Section 2, we introduce the terminology and set forth the panel decomposition. In Section 3 an empirical analysis is carried out to assess and compare the STC and other combinations. Section 4 presents the results of a bootstrap experiment. Finally, Section 5 concludes and puts forward some lines for future research.

2. Panel decomposition and the STC approach

Consider two periods: A *panel period* from 1 to T_1 and an *out-of-sample period* from $T_1 + 1$ to $T_2 - h$, where h denotes a forecast horizon. In the first one, we are *given* forecasts from J forecast units. The second one is then used to

compute and assess a forecast combination. In the STC approach, we first *split* the $T_1 \times J$ forecasts into M panels with $(T_1 / M) \times J$ forecasts each one, $m \in \{1, \dots, M\}$. Then, we choose J weights, $j \in \{1, \dots, J\}$, to compute the STC combination for that m .

Table 1. Panel decomposition in fixed and aggregate effects for panel m , h -steps-ahead forecasts

Panel m	1	2	...	J	$\overline{y}_{\bullet, t_1}^{(m)}$	Aggregate
1	$\hat{y}_{1, 1 0}^{(m)}$	$\hat{y}_{2, 1 0}^{(m)}$...	$\hat{y}_{J, 1 0}^{(m)}$	$\overline{y}_{\bullet, 1}^{(m)}$	$\left(\overline{y}_{\bullet, 1}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$
2	$\hat{y}_{1, 2 1}^{(m)}$	$\hat{y}_{2, 2 1}^{(m)}$...	$\hat{y}_{J, 2 1}^{(m)}$	$\overline{y}_{\bullet, 2}^{(m)}$	$\left(\overline{y}_{\bullet, 2}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$
...
T_1	$\hat{y}_{1, T_1 T_1-1}^{(m)}$	$\hat{y}_{2, T_1 T_1-1}^{(m)}$...	$\hat{y}_{J, T_1 T_1-1}^{(m)}$	$\overline{y}_{\bullet, T_1}^{(m)}$	$\left(\overline{y}_{\bullet, T_1}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$
$\overline{y}_{j\bullet}^{(m)}$	$\overline{y}_{1\bullet}^{(m)}$	$\overline{y}_{2\bullet}^{(m)}$...	$\overline{y}_{J\bullet}^{(m)}$	$\overline{\overline{y}}_{\bullet\bullet}^{(m)}$	
Fixed	$\left(\overline{y}_{1\bullet}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$	$\left(\overline{y}_{2\bullet}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$...	$\left(\overline{y}_{J\bullet}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$		

For panel m , let $\overline{y}_{j\bullet}^{(m)}$ be the j -column average of (T_1 / M) observations; $\overline{y}_{\bullet, t_1}^{(m)}$, the t_1 -row average of J observations, $t_1 = 1, \dots, (T_1 / M)$; and $\overline{\overline{y}}_{\bullet\bullet}^{(m)}$, the overall average. For instance, with monthly data from 2001.01 to 2010.12, $T_1 = 120$ and $M = 12$. If $J = 4$, then in the March panel, t_1 goes from 2001.03 to 2010.03, $\hat{y}_{2, t_1=2005}^{(3)}$ is the forecast of a March in year 2005 given by the 2nd forecast units, $\overline{y}_{2\bullet}^{(3)} = \frac{1}{10} \sum_{t_1=2001.03}^{2010.03} \hat{y}_{2, t_1}^{(3)}$, $\overline{y}_{\bullet, t_1=2005}^{(3)} = \frac{1}{4} \sum_{j=1}^4 \hat{y}_{j, t_1=2005}^{(3)}$, and

$$\overline{\overline{y}}_{\bullet\bullet}^{(3)} = \frac{1}{4} \sum_{j=1}^4 \overline{y}_{j\bullet}^{(3)} = \frac{1}{10} \sum_{t_1=2001.03}^{2010.03} \overline{y}_{\bullet, t_1}^{(3)} = \frac{1}{40} \sum_{t_1=2001.03}^{2010.03} \sum_{j=1}^4 \hat{y}_{j, t_1}^{(3)}.$$

2.1 Fixed, aggregate, and residual effects

We can always decompose the panel observations into three components: A *fixed* effect that captures intrinsic characteristics; an *aggregate* effect that captures common dynamics; and a residual, *idiosyncratic* error term. For panel m , consider the following identity,

$$\hat{y}_{j, t_1}^{(m)} \equiv \overline{y}_{j\bullet}^{(m)} + \left(\overline{y}_{\bullet, t_1}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right) + r_{j, t_1}^{(m)} = \overline{y}_{j\bullet}^{(m)} + \frac{1}{J} \sum_{j=1}^J \left(\hat{y}_{j, t_1}^{(m)} - \overline{y}_{j\bullet}^{(m)} \right) + r_{j, t_1}^{(m)}$$

where

- (i) $\overline{y}_{j\bullet}^{(m)} = \frac{1}{T_1 / M} \sum_{t_1=1}^{T_1 / M} \hat{y}_{j, t_1}^{(m)}$ is the j -fixed effect;
- (ii) $\overline{y}_{\bullet, t_1}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)}$ is the t_1 -aggregate effect;
- (iii) $r_{j, t_1}^{(m)}$ is the (j, t_1) -residual term.

Then,

$$V^{(m)} = \sum_{t_1=1}^{T_1 / M} \sum_{j=1}^J \left(\hat{y}_{j, t_1}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2 \equiv V_F^{(m)} + V_A^{(m)} + V_R^{(m)}$$

is, by construction, the sum of three orthogonal components:

$$V_F^{(m)} = \sum_{j=1}^J \left(\overline{y}_{j\bullet}^{(m)} - \overline{\overline{y}}_{\bullet\bullet}^{(m)} \right)^2$$

$$V_A^{(m)} = \sum_{t_1=1}^{T_1/M} \left(\bar{y}_{\bullet, t_1}^{(m)} - \bar{\bar{y}}_{\bullet\bullet}^{(m)} \right)^2$$

$$V_R^{(m)} = \sum_{t_1=1}^{T_1/M} \sum_{j=1}^J r_{j, t_1}^{(m)2}$$

In particular,

(i) When the ratio $V_A^{(m)} / V_R^{(m)}$ is large, we would expect a combination to capture better the *common* dynamics of forecasts;

(ii) The proportionate j -fixed effect,

$$w_{j, T_1}^{(m), STC} := \left(\bar{y}_{j\bullet}^{(m)} - \bar{\bar{y}}_{\bullet\bullet}^{(m)} \right)^2 / \sum_{k=1}^J \left(\bar{y}_{k\bullet}^{(m)} - \bar{\bar{y}}_{\bullet\bullet}^{(m)} \right)^2$$

defines our panel- m , sum-one STC weights $j \in \{1, \dots, J\}$ that form the combination

$$\hat{y}_{T_1+m|T_1}^{(m), STC} = w_{1, T_1}^{(m), STC} \hat{y}_{1, T_1+m|T_1}^{(m)} + w_{2, T_1}^{(m), STC} \hat{y}_{2, T_1+m|T_1}^{(m)} + \dots + w_{J, T_1}^{(m), STC} \hat{y}_{J, T_1+m|T_1}^{(m)}$$

outside the sample, where $\hat{y}_{j, T_1+m|T_1}^{(m)}$ denotes the (T_1+m) -step ahead forecast made at T_1 by the forecast unit j on panel m .

It is important to note that:

1.- The larger the distance between the mean of the predictions of a model and the overall mean, the larger the weight; i.e., the more different the model, the more informative its fixed effect is.

2.- We do not use the *actual* value of the target variable to compute the weights.

3.- Real *out-of-sample* forecasting is done;

4.- Weights *vary* for every forecast unit and panel.

5.- Only the restriction that the weights sum to unity is used.

6.- The procedure is valid for any forecast *horizon*.

7.- If an outlier or break is in the panel period, the STC weights will adapt as soon as the forecasts do.

8.- If an outlier or break is in the out-of-sample period, nothing can be done. The combination will, in general, be biased. However, as pointed out by Elliot & Timmermann (2005) or Guidolin & Fangzou Na (2007), *time-varying* weights will usually capture the presence of structural shifts in the target variable.

2.2 Sum-one precision (SOP) weights

To compare the accuracy of our STC procedure, we use the varying-weight SOP combination as a benchmark,

$$\hat{y}_{T_1+m|T_1}^{(m), SOP} = w_{1, T_1}^{(m), SOP} \hat{y}_{1, T_1+m|T_1}^{(m)} + w_{2, T_1}^{(m), SOP} \hat{y}_{2, T_1+m|T_1}^{(m)} + \dots + w_{J, T_1}^{(m), SOP} \hat{y}_{J, T_1+m|T_1}^{(m)}$$

with

$$w_{j, T_1}^{(m), SOP} := \frac{\xi_{j, T_1}^{(m)}}{\xi_{1, T_1}^{(m)} + \dots + \xi_{J, T_1}^{(m)}}$$

where $\xi_{j, T_1}^{(m)}$ is the inverse of the average standard error of the j -forecast unit for panel m .

3. Application

3.1 Data collection, variables, and forecasting models

For this application, we have chosen the variables and models considered in García-Ferrer, de Juan, & Poncela (2006). There are three reasons to do so. First, the variables may be freely downloaded from *Dirección General de Tráfico* (Note 1) webpage, so we can easily update the sample at any time; second, these time series cover a long period of monthly data from 1975 to 2010 (432 monthly observations); third, the models are already identified and easy to implement (Note 2).

The variables are: *ACC* (total number of traffic accidents in Spain), *ACCR* (road traffic accidents), *ACCU* (urban traffic accidents), *FAT* (fatalities in the following 24 hours after the accident), *FATR* (fatalities after a road accident), *INJ* (injured passengers), *INJR* (injured passengers in road traffic accidents), and *INJU* (injured passengers in urban traffic accidents). Therefore, the number of target variables is $N = 8$. The evolution of the variables in Figure 1 shows a changing seasonal behavior over time, particularly at the end of the sample.

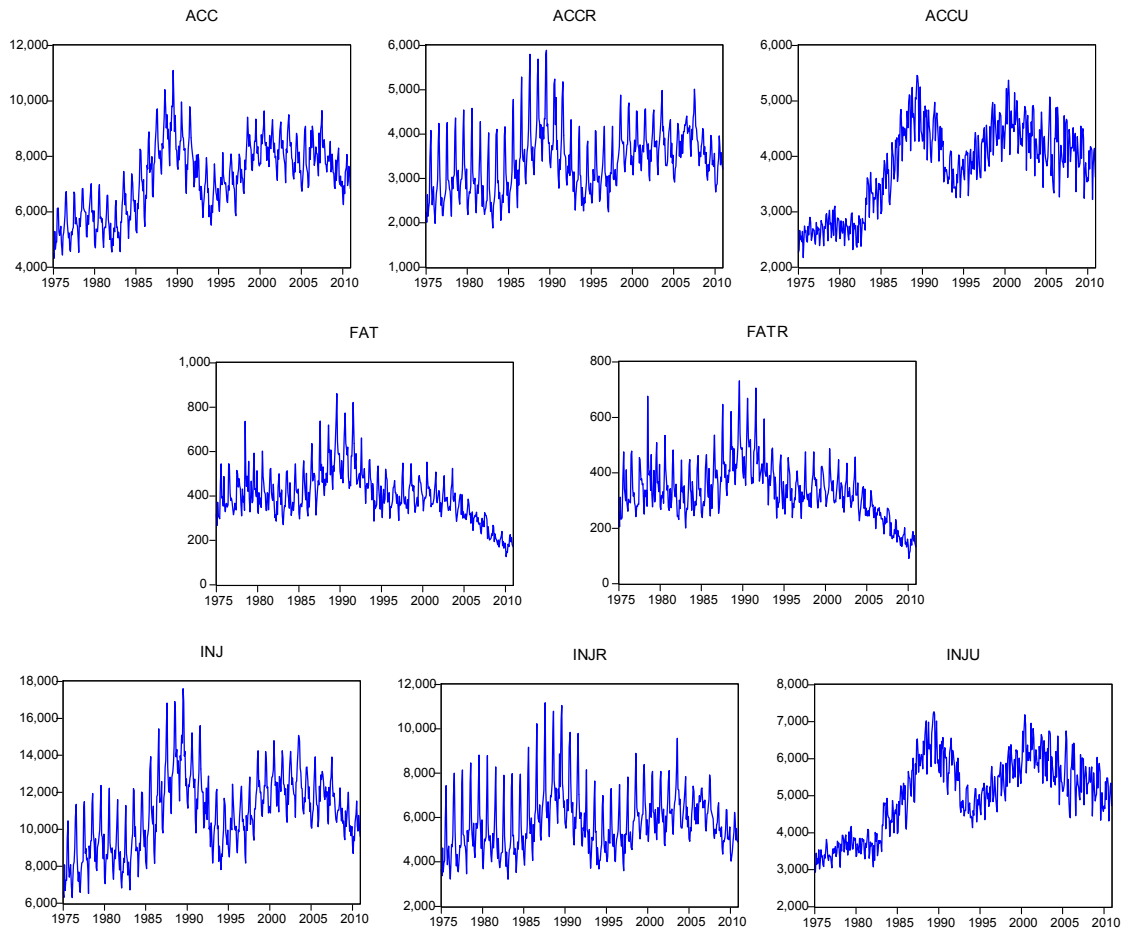


Figure 1. Evolution of the variables in the application. Sample: 1975M1-2010M12

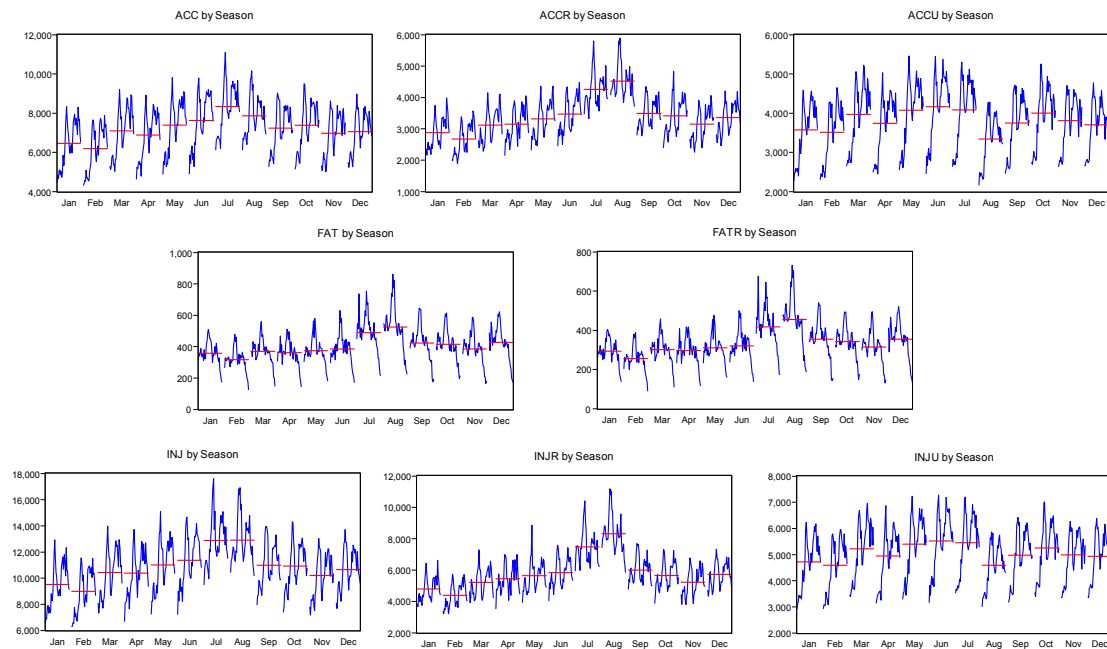


Figure 2. Means by season for each one of the variables

This fact can also be appreciated in Figure 2 that shows the average by seasons of each variable. In particular, during the summer season, August was by far the most dangerous month in terms of number of accidents, fatalities and injured passengers. The changing seasonality implies in our case that the average data for summer months gradually shrinks because accidents, fatalities, and injured passengers are spread out over three months (July, August, and September) instead of just only one (August). Figure 2 shows a decreasing August and a jumping July and September. From the beginning of the sample, August data are the highest, while along the end there are cases in which July's are higher.

Models for each variable are explained in García-Ferrer *et al.* (2006) and summarized in Table 2. Model 1, $ARIMA(0, 1, 1) \times ARIMA(0, 1, 1)_{s=12}$, is a benchmark with some interventions that pick up several additive outliers. The other five are transfer function models. Consistently estimated filters in models 2, 3, and 4 are *not nested* in models 5 and 6. Model forecasts contain different information that allows for model combination with *no encompassing*. The potential problem in a situation where some (or all) of the models we combine are nested is that the innovations from nested models can exhibit high cross-sectional dependence, thus preventing a weak LLN from holding (Chamberlain & Rothschild, 1983; Stock & Watson, 2002) unless the smallest model is correctly specified. *Given a correct specification of the smallest model*, its population forecast will be the conditional expectation, making the bias and residual effect zero, as well as those models that nest it. Any model that nests the smallest, correct one will have irrelevant estimated parameters that will converge to zero in probability (Note 3).

Table 2. (*b*, *r*, *s*) identified orders in the estimated transfer function models for the variables included in the application

	Model 2	Model 3	Model 4	Model 5	Model 6
ACC	$x_{1,t} : (1, 0, 0)$	$x_{2,t} : (1, 0, 1)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 1)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 1)$ $x_{3,t} : (0, 0, 0)$
ACCR	$x_{1,t} : (1, 0, 1)$	$x_{2,t} : (1, 0, 0)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$ $x_{3,t} : (0, 0, 0)$
ACCU	$x_{1,t} : (0, 0, 0)$	$x_{2,t} : (1, 0, 1)$	$x_{3,t} : (1, 0, 0)$ $x_{4,t} : (0, 0, 0)$	$x_{1,t} : (0, 0, 0)$ $x_{2,t} : (1, 0, 1)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 1)$ $x_{4,t} : (0, 0, 0)$
FAT	$x_{1,t} : (1, 0, 0)$	$x_{2,t} : (1, 0, 0)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 1)$ $x_{3,t} : (0, 0, 0)$
FATR	$x_{1,t} : (1, 0, 0)$	$x_{2,t} : (1, 0, 1)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 1)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$ $x_{3,t} : (0, 0, 0)$ $x_{4,t} : (2, 0, 0)$
INJ	$x_{1,t} : (1, 0, 1)$	$x_{2,t} : (1, 0, 0)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$ $x_{3,t} : (0, 0, 0)$
INJR	$x_{1,t} : (1, 0, 1)$	$x_{2,t} : (1, 0, 0)$	$x_{3,t} : (0, 0, 0)$ $x_{4,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 1)$ $x_{2,t} : (1, 0, 0)$	$x_{1,t} : (1, 0, 0)$ $x_{2,t} : (1, 0, 0)$ $x_{3,t} : (0, 0, 0)$
INJU	$x_{1,t} : (1, 0, 1)$	$x_{2,t} : (1, 0, 1)$	$x_{3,t} : (1, 0, 0)$ $x_{4,t} : (0, 0, 0)$	$x_{1,t} : (0, 0, 0)$ $x_{2,t} : (1, 0, 1)$	$x_{1,t} : (0, 0, 0)$ $x_{2,t} : (1, 0, 1)$ $x_{3,t} : (1, 0, 0)$

The triplet (b, r, s) correspond to the orders of the polynomials in B in the usual definition of a transfer function model, that is: $y_t = \frac{\omega(B)}{\delta(B)} B^b x_{i,t} + N_t$, with $\omega(B) = (\omega_0 + \omega_1 B + \dots + \omega_s B^s)$, $\delta(B) = (1 - \delta_1 B - \dots - \delta_r B^r)$ and N_t is the noise of the model usually identified as $ARIMA(0, 1, 1) \times (0, 1, 1)_{12}$. The variables $x_{i,t}$ are: $x_{1,t} = IPI$, $x_{2,t} = NUVE$, $x_{3,t} = CGAS$ and $x_{4,t} = CDIESEL$. All the variables in the transfer function models are in logs to achieve the stationarity.

3.2 The Sample

(i) First subsample (Note 4): The estimation period is from January, 1975 to December, 1985 with 132 observations (11 years of monthly data);

(ii) Second subsample: The panel period is from January, 1986 to December, 2005 with 240 observations (20 years of monthly data). In this subsample, we

(a) Recursively generate one-, six-, and twelve-step-ahead forecasts for each model ($j = 1, \dots, 6$) and variable ($i \in \{1, \dots, N = 8\}$);

(b) *Split into panels*: For each forecast horizon ($h = 1, 6, 12$), we classify by months the time dimension of the forecast panel to form a twelve-panel decomposition with 120 (20×6) month-pegged observations each one. We then calculate the STC and SOP weights for each model.

(iii) Third subsample: A first out-of-sample period is from January, 2006 to December, 2006. This subsample is used to assess three combination models; two of them based on panel-decomposition weights, and a third one given by a benchmark average of forecasts. We re-do this forecasting exercise for years 2007-2010.

3.3 Results and Discussion

In evaluating the forecasting accuracy of a forecast combination over the out-of-sample period, no single error measure captures the distributional features of the forecast errors when summarized across data series (Hyndman & Koehler, 2006). Hence, we consider several standard accuracy measures commonly used in evaluating the accuracy of forecasts (Note 5). In general, we find that APE and FGR are good for *medium-term* forecasting while the others are useful for *short-term* forecasting. Beside these global measures, we also considered monthly percentage prediction errors (PPE).

The data evolve in a non-stationary manner and exhibit obvious variations in their seasonal patterns. Therefore, we expect larger forecast errors concentrated on those months, where seasonality is more variable, that may be counterbalanced with longer forecast horizons, thus making the use of RMSE and MAPE not only inappropriate but also misleading. Also, the forecast ability of the models is ambiguous. Models M1, M2 and M4 are better in terms of PPE and global measures, but none of them is best for all months and all forecast horizons (Note 6).

3.3.1 Accuracy measures for combinations

Tables in this section summarize the three combinations used in the Application. We show the results for *ACC* only (Note 7). Panel models make possible to condition the data generation process of a time series entirely on the season; thus, appreciably reducing periodically varying forecast errors.

Table 3 shows the number of beats across variables and monthly PPE. STC is the best while the Average model is the worst. Only in 2008, $h = 1$, the latter beats SOP. Varying-weight combinations (STC and SOP) outperform, in general, the fixed-weight one (Average model). We only find two matches in Average and SOP (2006, $h = 6$ and $h = 12$).

Table 3. Number of beats across variables and percentage prediction errors (%) - Forecast Combinations

		$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$	Total
2006	$h = 1$	27	44	26	97
	$h = 6$	24	50	24	98
	$h = 12$	25	48	25	98
	Total	76 (25.9)	142 (48.5)	75 (25.6)	293
2007	$h = 1$	29	45	25	99
	$h = 6$	34	39	26	99
	$h = 12$	27	45	24	96
	Total	90 (30.6)	129 (43.9)	75 (25.5)	294
2008	$h = 1$	29	39	30	98
	$h = 6$	41	42	14	97
	$h = 12$	34	49	13	96
	Total	104 (35.7)	130 (44.7)	57 (19.6)	291
2009	$h = 1$	42	43	12	97
	$h = 6$	34	45	30	109
	$h = 12$	32	47	20	99
	Total	108 (35.4)	135 (44.3)	62 (20.3)	305
2010	$h = 1$	32	45	22	99
	$h = 6$	33	48	16	97
	$h = 12$	21	50	23	94
	Total	86 (29.7)	143 (49.3)	61 (21.0)	290
Total 06+07+08+09+10		464 (31.5)	679 (46.1)	330 (22.4)	1473

$\hat{y}_M^{(P)}$ = Sum-one weights, precision combination; $\hat{y}_M^{(STC)}$ = STC combination; $\bar{\hat{y}}_{T+l}$ = Model Average combination.

Table 4 shows the number of beats across variables and accuracy measures. Again, STC is the best (with the only exception of 2007). A fact already found in the previous section is that, for all forecast horizons, SOP outperforms STC in 2007, the first complete year in which a new traffic regulation was in force (Note 8). Taken together all the years, STC outperforms SOP by more than 10 percentage points (44.9% vs 33.5%). Notice that the fixed-weight is never better than any of the varying-weight combinations.

Table 4. Number of beats across variables and accuracy measures (%)

		$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$	Total
2006	$h = 1$	15	42	14	71
	$h = 6$	20	33	23	76
	$h = 12$	15	40	23	78
	Total	50 (22.2)	115 (46.0)	60 (26.7)	225
2007	$h = 1$	37	23	20	80
	$h = 6$	38	21	25	85
	$h = 12$	34	26	20	80
	Total	109 (44.5)	70 (28.6)	65 (26.5)	245
2008	$h = 1$	18	38	5	61
	$h = 6$	32	34	8	74
	$h = 12$	21	41	7	69
	Total	71 (34.8)	113 (55.4)	20 (9.8)	204
2009	$h = 1$	29	35	18	82
	$h = 6$	20	22	17	59
	$h = 12$	28	31	21	80
	Total	77 (34.8)	88 (39.8)	56 (25.3)	221
2010	$h = 1$	21	37	14	72
	$h = 6$	18	38	12	68
	$h = 12$	23	34	11	68
	Total	62 (29.8)	109 (52.4)	37 (17.8)	208
Total 06+07+08+09+10		369 (33.5)	495 (44.9)	238 (21.6)	1112

$\hat{y}_M^{(P)}$ = Sum-one weights precision combination; $\hat{y}_M^{(STC)}$ = STC combination; \hat{y}_{T+l} = Model Average combination.

Figure 3 shows the STC weights for different combination periods (2006 to 2010). This figure helps to appreciate its evolving weights for each month and year.

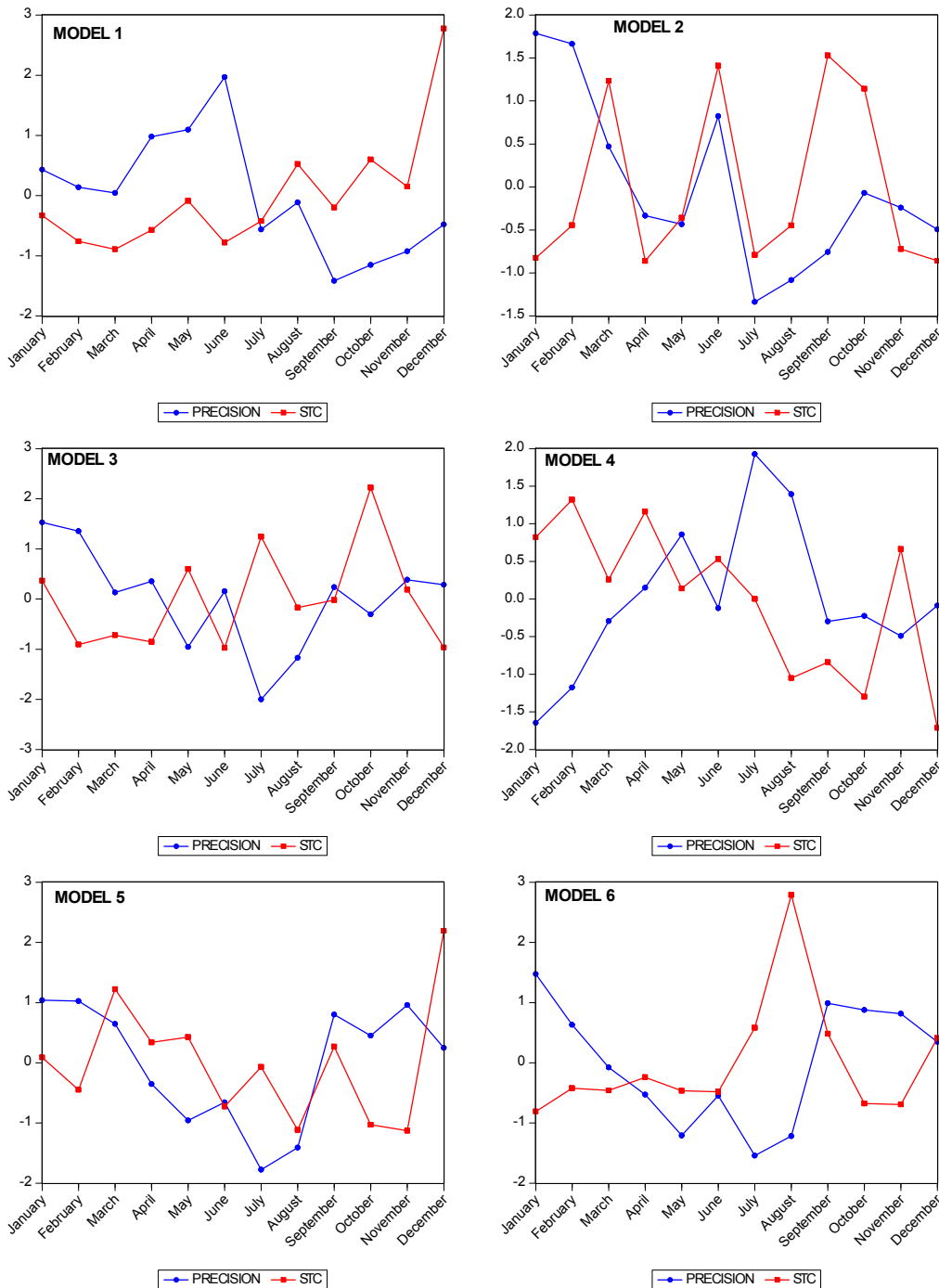


Figure 3. Precision and STC weights, 2006, ACC variable, h=1

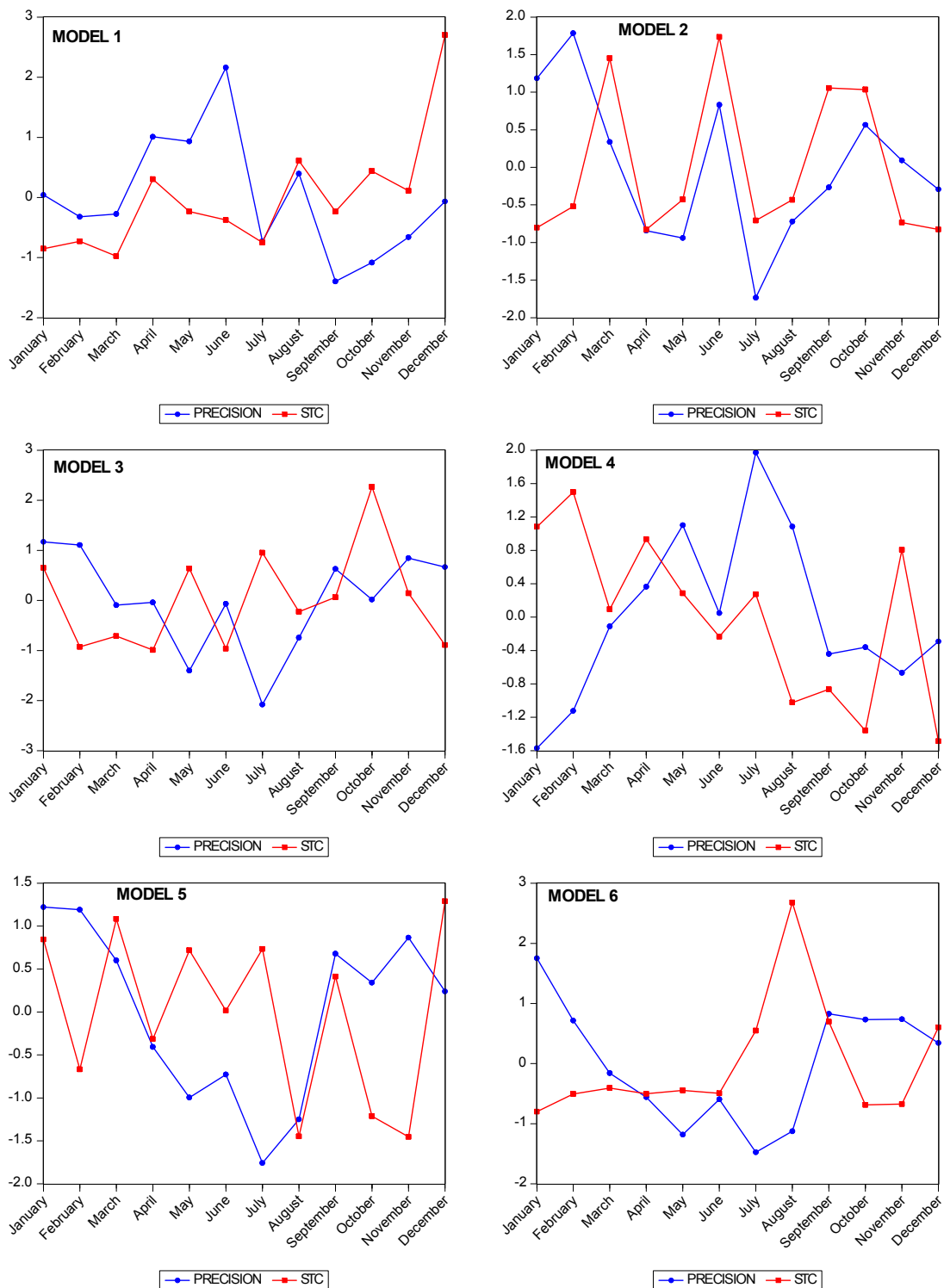


Figure 3. Precision and STC weights, 2007, ACC variable, h=1

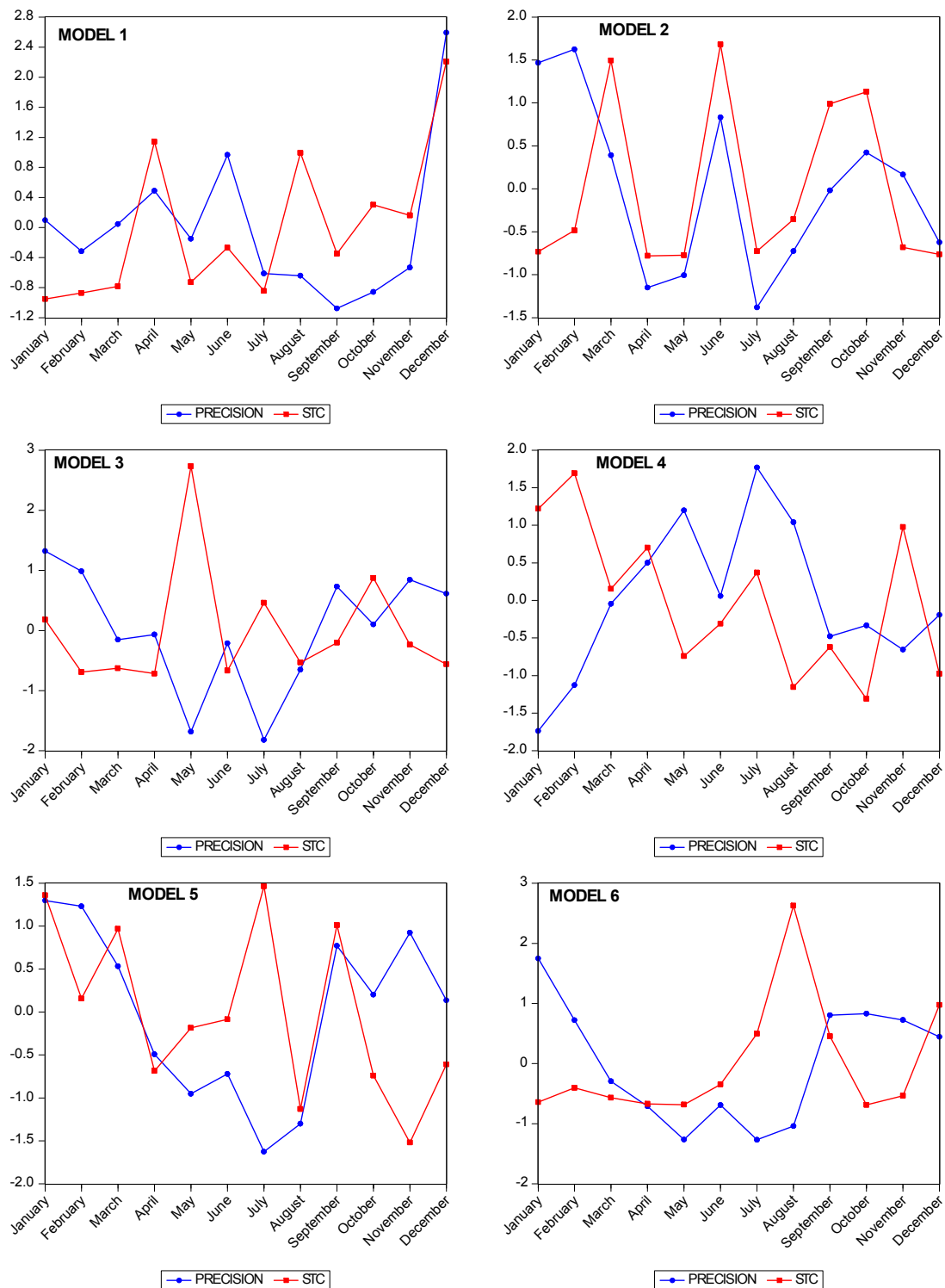
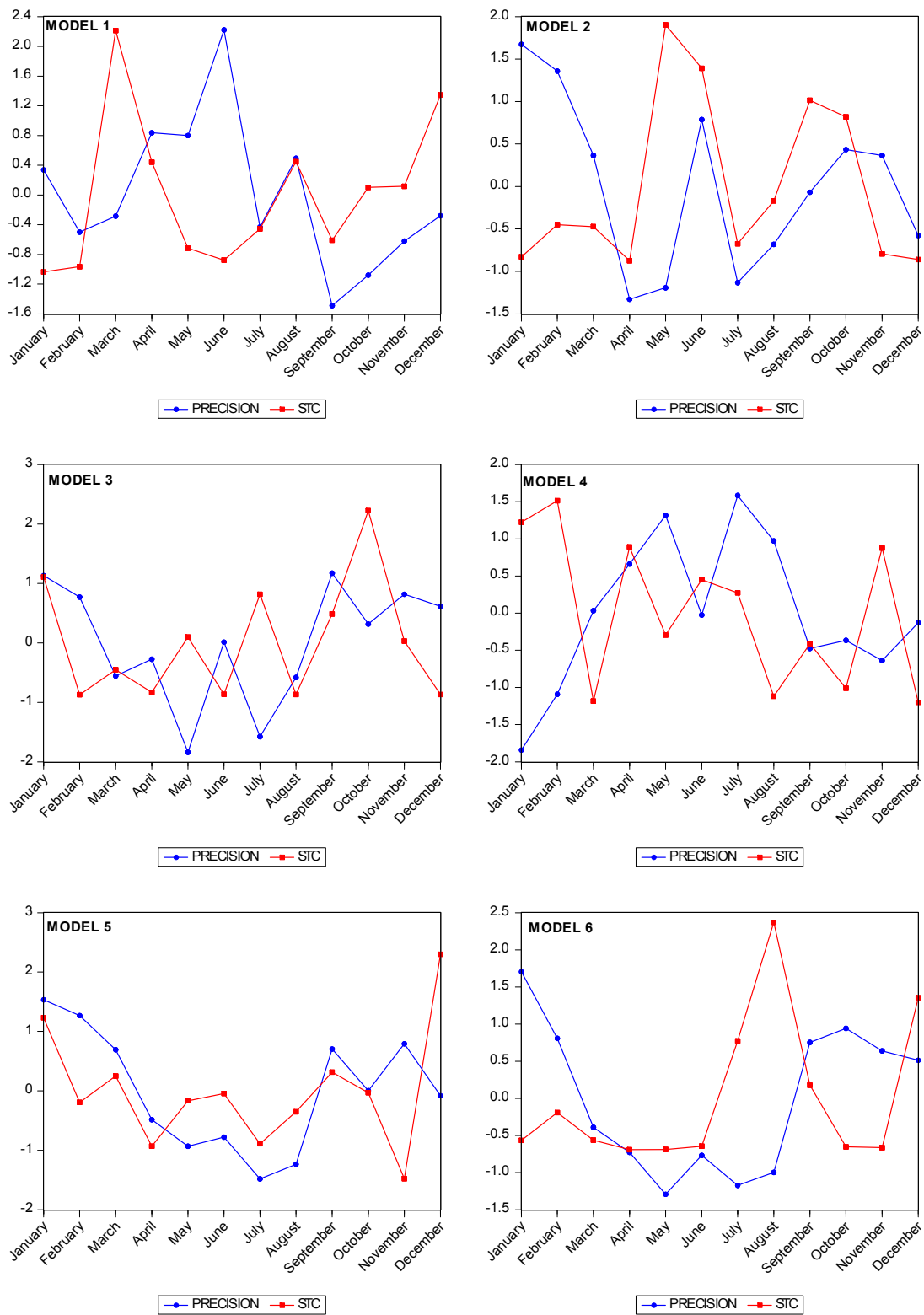


Figure 3. Precision and STC weights, 2008, ACC variable, h=1

Figure 3. Precision and STC weights, 2009, ACC variable, $h=1$

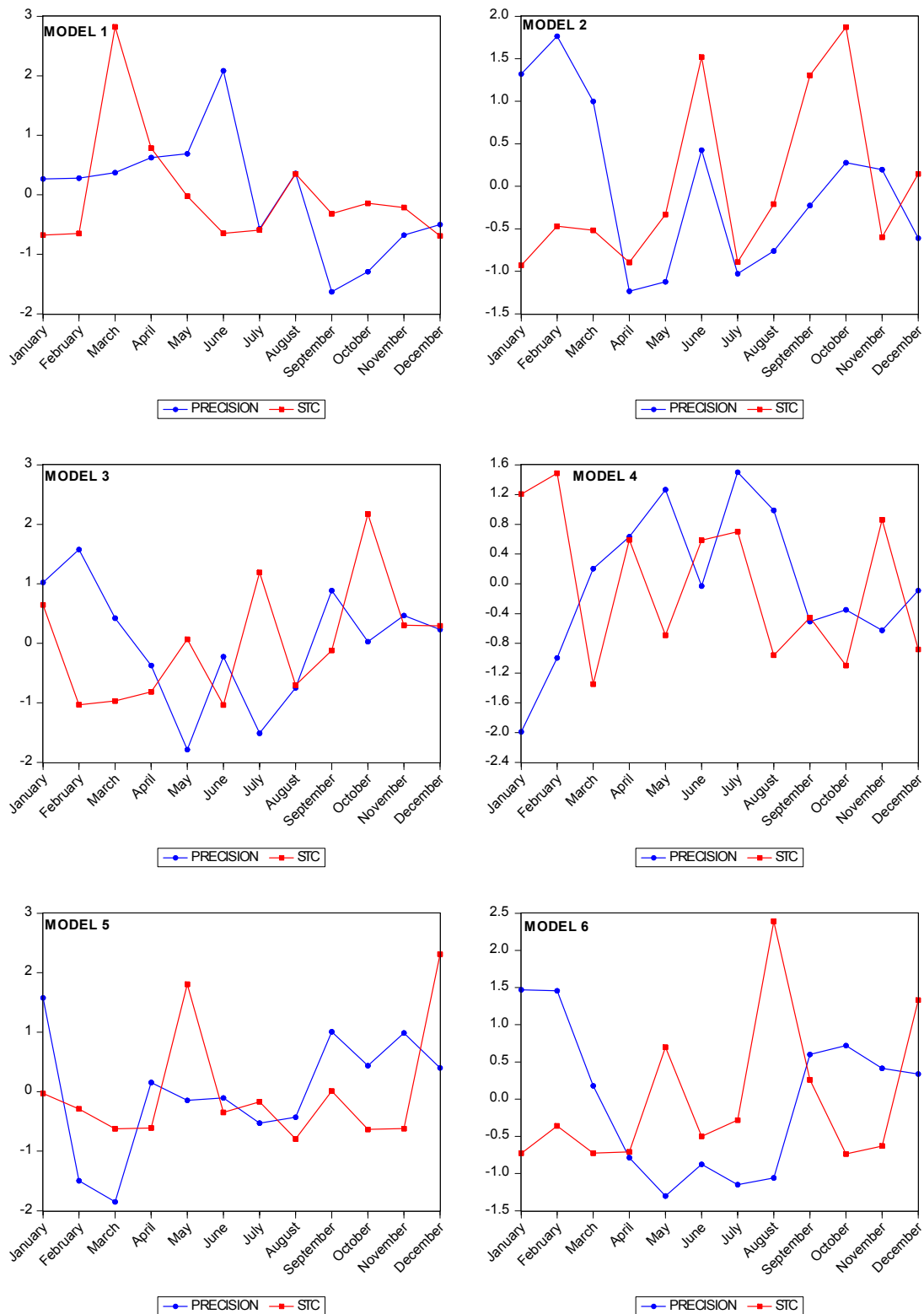


Figure 3. Precision and STC weights, 2010, ACC variable, $h=1$

Finally, Tables 5 and 6 summarize the results by accuracy criteria. STC always wins for $h = 1$ and $h = 12$. For $h = 6$, STC wins in five accuracy criteria, and SOP in two (MAPE and MdAPE). Looking at the total number of beats, STC

is always the winner while the fixed-weight is always the worst. Hence, varying-weight combinations always outperform the fixed-weight one.

These results confirm the need of periodic-weight combinations of forecasts. As a by-product, we found no definitive criterion for choosing a specific combination. Notice, however, that for RMSPE and sMAPE, there are so many matches that it makes difficult to distinguish among combinations.

Table 5. Number of beats by accuracy criteria and forecast horizons

	<i>h</i> = 1			<i>h</i> = 6			<i>h</i> = 12		
	$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$	$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$	$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$
<i>APE</i>	11	22	10	16	<i>17</i>	8	13	<i>20</i>	12
<i>FGR</i>	11	22	9	16	<i>17</i>	9	13	<i>20</i>	12
<i>RMSE</i>	21	26	17	20	22	16	18	27	13
<i>MAPE</i>	22	27	10	23	22	14	21	27	12
<i>MdAPE</i>	13	23	6	25	23	12	23	25	9
<i>sMAPE</i>	23	27	12	18	24	16	16	27	17
<i>sMdAPE</i>	18	25	7	22	23	12	17	26	8
<i>TOTAL</i>	119	172	71	140	148	87	121	172	83
%	32.9	47.5	19.6	37.3	39.5	23.2	32.2	45.7	22.1

$\hat{y}_M^{(P)}$ = Sum-one weights precision combination; $\hat{y}_M^{(STC)}$ = STC combination; $\bar{\hat{y}}_{T+l}$ = Model Average combination. Winners in italics, while highest percentages in bold.

Table 6. Summary results by accuracy criteria: Number of beats (%)

	$\hat{y}_M^{(P)}$	$\hat{y}_M^{(STC)}$	$\bar{\hat{y}}_{T+l}$
<i>APE</i>	40	<i>59</i>	32
<i>FGR</i>	40	<i>59</i>	31
<i>RMSPE</i>	59	85	46
<i>MAPE</i>	63	76	34
<i>MdAPE</i>	47	74	25
<i>sMAPE</i>	67	78	46
<i>sMdAPE</i>	53	74	24
<i>TOTAL</i>	369 (33.1)	505 (45.4)	238 (21.4)

$\hat{y}_M^{(P)}$ = Sum-one weights precision combination; $\hat{y}_M^{(STC)}$ = STC combination; $\bar{\hat{y}}_{T+l}$ = Model Average combination. Winners in italics, while highest percentages in bold.

4. Bootstrap Analysis

In this section, we are interested in the *small-(J, T_i)* behavior of the STC *weights* associated to the models corresponding to each month in our panel decomposition. Approximate properties of the *theoretical* distribution of model weights are untrustworthy because the panel size (years 1986 to 2009) and the number of models (6) are small. The alternative is to estimate the properties we require from *simulated* panel, model-forecast datasets. A *bootstrap* analysis based on the *out-of-sample* forecast errors usually requires the *forecast errors* to be serially uncorrelated, as would typically be the case in forecasting models at horizon $h = 1$. At longer horizons, the forecast errors will not be *iid* and the standard bootstrap method may not be valid. A possible solution to this problem is to construct forecast errors and recover the underlying white noise errors by fitting an *MA(h-1)* process to the sequence of *h*-step-ahead forecast errors. This allows the construction of bootstrap samples by first resampling the serially uncorrelated white noise residuals using standard bootstrap methods, and then constructing bootstrap replicates of the *h*-month-ahead forecast errors from the implied moving averages. Another solution is to fit the forecasting models on non-overlapping observations and proceed as for $h = 1$. This approach is simple, but may involve a considerable reduction in estimation precision. Remember, however, that the STC combination weights are *not* based on *forecast errors*.

In this study, we have generated bootstrap samples for general *stationary* data according to Politis & Romano (1994) and computed the bootstrap estimate of the standard error and bias of the STC weights (Note 9). For consistency, the (mean) block length should grow with the panel size at an appropriate rate. A default growth rate proportional to the third root of the panel size is used. This rate is "optimal" under certain conditions. However, in general the growth rate depends on the specific properties of the data generation process. A default value for the proportionality constant has been determined by a Monte Carlo simulation using a Gaussian *AR(1)* process [*AR(1)*-parameter of 0.5, 500 observations] and the constant has been chosen such that the MSE for the bootstrap estimate of the variance of the empirical mean is minimized. All the computing has been carried out with R (Note 10). Once the bootstrap samples for each model are obtained, we apply the STC procedure to get replicates of their associated STC weights. Then, by linearly applying the bootstrapped panel STC weights to the out-of-sample model forecasts, we obtain the bootstrapped bias and SE of the out-of-sample STC forecast.

As an example, Table 7 refers to the variable *ACC* (Note 11). It shows, for $h = 6$ and $h=12$, the bootstrapped *squared bias* of each model's weight as a percentage of its bootstrapped MSE (six first columns). The last two columns compare the performance of the out-of-sample (year 2010) bootstrapped absolute forecast errors of the STC and Average combinations using the bootstrap replicates from the bootstrapped panels, taken the actual value of *ACC* and model forecasts in the out-of-sample period as given (Note 12). In general, model weight biases are statistically insignificant and slightly smaller for $h = 6$ as expected. The bootstrapped weighting scheme differs depending on the month, but there is no evidence of a consistent superiority of any model's weight in all months. On the other hand, bootstrapped combinations reflect a low August with a jumping afterward also found in the Application. All this makes it even more attractive the use of combinations. In particular, STC has been able to properly weight each model in such a way that systematically outperforms the average of them. The Average has consistently higher bootstrapped SEs than those of STC's. This result remain statistically unchanged among different variables, months, and forecast horizons, and it is consistent with other Monte Carlo studies where the number of models is sufficiently small compared with the number of observations in the training sample (see, for instance, Issler & Lima, Table 3). Although the optimal-weighted forecast and a bias-corrected average forecast are *asymptotically* equivalent (and this is one way of solving the combination puzzle), what we have found here is that, in *small samples* with a *small number of forecast units*, the STC with panel-varying weights fares better than an Average. The greater accuracy of the forecasts based on STC weights relative to those based on fixed ones may reflect a good small-sample approximation of optimal population weights by panel-varying weights. This may be considered as another look to solving the *puzzle*.

Table 7. Bias and Stantard Error (SE) of the STC and Average (AVE) forecast errors for 2010 ACC variable - Bootstrap Simulation Results

		h=1		h=6		h=12	
		Bias	SE	Bias	SE	Bias	SE
January	$\hat{y}_M^{(STC)}$	-7.39	7.41	6.10	7.85	8.55	10.21
	$\bar{\hat{y}}_{T+l}$	-1.94	159.50	-2.46	169.84	-5.75	170.86
February	$\hat{y}_M^{(STC)}$	8.23	9.65	3.53	8.17	1.86	9.53
	$\bar{\hat{y}}_{T+l}$	5.81	163.10	5.21	171.30	4.69	178.80
March	$\hat{y}_M^{(STC)}$	4.15	10.33	0.13	14.50	-8.90	15.86
	$\bar{\hat{y}}_{T+l}$	4.96	183.77	3.99	198.19	8.29	195.51
April	$\hat{y}_M^{(STC)}$	3.90	14.77	1.79	14.73	2.00	15.36
	$\bar{\hat{y}}_{T+l}$	2.08	186.71	-0.53	190.14	-0.39	186.21
May	$\hat{y}_M^{(STC)}$	-1.67	4.23	-8.88	7.10	-6.36	8.50
	$\bar{\hat{y}}_{T+l}$	-5.59	164.07	3.52	197.36	-1.06	172.81
June	$\hat{y}_M^{(STC)}$	-3.79	5.69	5.45	4.91	0.58	4.91
	$\bar{\hat{y}}_{T+l}$	-7.08	188.19	5.37	164.68	3.56	164.68
July	$\hat{y}_M^{(STC)}$	0.70	3.25	4.98	8.57	2.96	7.80
	$\bar{\hat{y}}_{T+l}$	-3.35	154.93	-6.75	160.45	0.75	164.70
August	$\hat{y}_M^{(STC)}$	-7.11	19.89	2.68	22.71	9.73	23.22
	$\bar{\hat{y}}_{T+l}$	-2.33	138.26	-0.60	144.25	1.66	140.69
September	$\hat{y}_M^{(STC)}$	0.17	14.44	5.45	60.03	1.85	14.80
	$\bar{\hat{y}}_{T+l}$	6.94	139.56	11.62	140.36	-2.76	142.38
October	$\hat{y}_M^{(STC)}$	25.31	71.88	4.37	8.86	-4.36	8.69
	$\bar{\hat{y}}_{T+l}$	60.22	101.57	41.00	121.27	28.45	114.15
November	$\hat{y}_M^{(STC)}$	-3.09	7.55	-1.37	6.43	-0.82	6.00
	$\bar{\hat{y}}_{T+l}$	18.01	145.21	35.89	132.11	35.65	130.12
December	$\hat{y}_M^{(STC)}$	5.90	6.02	1.73	2.98	1.41	2.59
	$\bar{\hat{y}}_{T+l}$	1.28	152.51	-5.28	155.93	-7.56	149.44

5. Conclusion

Relative forecast performance of forecast units may periodically evolve over time. Therefore, it is desirable to take into account their forecast *periodicity* when forming forecast combinations. When dealing with *small* samples and number of forecast units, using *panels* is an efficient way of pulling out the additional information provided by that periodicity in the data, in most cases by an *evolving seasonality* that may not be well captured by standard models with stationary seasonality. We compute weights from panels and keep them fixed in the out-of-sample period. Thus, the *forecast combination puzzle* associated with *estimated* weights is not an issue with our procedure. Empirical and *bootstrap* exercises illustrate the superiority of this method over fixed weight schemes. With monthly data for eight

traffic accident variables in Spain, we decompose the information given by six model forecasts into twelve panels and form two varying-weight forecast-combinations that have outperformed, according to seven accuracy criteria, one-, six, and twelve-step-ahead out-of-sample model forecasting as well as a standard fixed-weight combination such as the average of them. In particular, our Split-then-Combine combination is the best choice. Panel decompositions may also be used to form panel forecast combinations after a previous individual-model selection is carried out (Combination-after-Selection), possibly resulting in different models, or number of them, for different periods. When the number J of forecast units is large, this approach would have the expected effect of acting as panel-based, dimension-reduction procedure that may complement others such as period-based principal components or seasonal dynamic factor analysis. These are extensions that we are currently investigating.

References

- Basawa, I.V. & Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, 22, 651-63. <http://dx.doi.org/10.1111/1467-9892:00245>
- Bates, J.M. & Granger, C.W.J. (1969). The combination of forecasts. *Operations Research Quarterly*, 20, 451-68. <http://dx.doi.org/10.1057/jors.1969.103>
- Boswijk H.P., & Franses, P.H. (1996). Unit roots in periodic autoregressions. *Journal of Time Series Analysis* 17, 221-246. <http://dx.doi.org/10.1111/j.1467-9892.tb00274.x>
- Canty, A. & Ripley, B. (2011). boot: Bootstrap R (S-Plus) Functions. <http://CRAN.R-project.org/package=boot>.
- Chamberlain, G. & Rothschild, M. (1983): Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51, 1281-1304. <http://dx.doi.org/10.2307/1912275>
- Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2. <http://dx.doi.org/10.1017/CBO9780511802843>
- Deutsch, M. Granger, C.W.J., & Terasvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting*, 10, 47-57. [http://dx.doi.org/10.1016/0169-2070\(94\)90049-3](http://dx.doi.org/10.1016/0169-2070(94)90049-3)
- Elliott, G. & Timmermann, A. (2005). Optimal forecast combination weights under regime switching. *International Economic Review*, 46, 1081-1102. <http://dx.doi.org/10.1111/j.1468-2354.2005.00361.x>
- Franq, C. Roy, R., & Saidi, A. (2011). Asymptotic properties of weighted least squares estimation in weak PARMA models. *Journal of Time Series Analysis*, 32, 699-723. <http://dx.doi.org/10.1111/j.1467.9892.2011.00728.x>
- Franses, P.H. & Paap, R. (2004). *Periodic Time Series Models*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/019924202X.001.0001>
- García-Ferrer, A., de Juan, A., Poncela, P. & Bujosa, M. (2004). Monthly forecasts of integrated public transport systems: The case of the Madrid metropolitan area. *Journal of Transportation and Statistics*, 7(1), 39-59.
- García-Ferrer, A., Bujosa, M., de Juan, A. & Poncela, P. (2006). Demand forecast and elasticities estimation of public transport. *Journal of Transport Economics and Policy*, 40(1), 45-67.
- García-Ferrer, A., de Juan, A. & Poncela, P. (2006). Forecasting traffic accidents using disaggregated data. *International Journal of Forecasting*, 22 (2), 203-22. <http://dx.doi.org/10.1016/j.ijforecast.2005.11.001>
- Granger, C.W.J. & Ramanathan, R. (1984). Improved methods of combining forecasting. *Journal of Forecasting*, 3, 197-204. <http://dx.doi.org/10.1002/for.3980030207>
- Guidolin, M. & Fangzhou Na, C. (2007). The economic and statistical value of forecast combinations under regime switching: An application to predictable US returns. *Working Paper 2006-059B*, Research Division, Federal Reserve Bank of St. Louis.
- Guidolin, M. & Timmermann, A. (2009). Forecasts of US short-term interest rates: A flexible forecast combination approach. *Journal of Econometrics*, 150, 297-311. <http://dx.doi.org/10.1016/j.jeconom.2008.12.004>
- Herwartz, H. (1999). Performance of periodic time series models in forecasting. *Empirical Economics*, 24, 271-301. <http://dx.doi.org/10.1007/s001810050055>
- Hoogerheide, L., Kleijn, R., Ravazzolo, R., van Dijk, H.K. & Verbeek, M.(2010). Forecast accuracy and economic gains from Bayesian model averaging using time varying weight. *Journal of Forecasting*, 29 (1-2), 251-269. <http://dx.doi.org/10.1002/for.1145>
- Hyndman, R.I., & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of*

- Forecasting*, 22, 679-88. <http://dx.doi.org/10.1016/j.ijforecast.2006.03.001>
- Issler, J.V. & Lima, L.R. (2009). A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics*, 152, 153-164. <http://dx.doi.org/10.1016/j.jeconom.2009.01.002>
- Jones, R. & Brelsford, W. (1967). Time series with periodic structure. *Biometrika*, 54, 403-8.
- Kim, J.H. (2009). BootPR: Bootstrap Prediction Intervals and Bias-Corrected Forecasting. <http://CRAN.R-project.org/package=BootPR>.
- LeSage, J.P. & Magura, M. (1992). A mixture-model approach to combining forecasts. *Journal of Business and Economic Statistics*, 10, 445-53. <http://dx.doi.org/10.1080/07350015.1992.10509920>
- Matas-Mir, A., Osborn, D. & Lombardi, J. (2008). The effect of seasonal adjustment on the properties of business cycle regimes. *Journal of Applied Econometrics*, 23(2), 257-278. <http://dx.doi.org/10.1002/jae.980>
- Novalés, A. & Flores de Frutos, R. (1997). Forecasting with periodic models: A comparison with time invariant coefficient models. *International Journal of Forecasting*, 13, 393-405. [http://dx.doi.org/10.1016/S0169-2070\(97\)00026-5](http://dx.doi.org/10.1016/S0169-2070(97)00026-5)
- Osborn, D. & Smith, J. (1989). The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business and Economic Statistics*, 7, 117-27. <http://dx.doi.org/10.1080/07350015.1989.10509719>
- Politis, D.N. & Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89, 1303--1313. <http://dx.doi.org/10.1080/01621459.1994.10476870>
- Poncela, P. Rodríguez, J., Sánchez-Mangas, R., & Senra, E. (2011). Forecast combination through dimension reduction techniques. *International Journal of Forecasting*, 27, 224-237. <http://dx.doi.org/10.1016/j.ijforecast.2010.01012>
- Ravazzolo, F., van Dijk, H.K. & Verbeek, M. (2007). Predictive gains from forecast combinations using time varying model weights. *Econometric Institute Report 2007-26*.
- Sánchez, I. (2008). Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, 24, 679-693. <http://dx.doi.org/10.1016/j.ijforecast.2008.08.008>
- Smith, J. & Wallis, K.K. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3), 331-355. <http://dx.doi.org/10.1111/j.1468-0084.2008.00541.x>
- Stock, J., & Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 460, 1167-1179. <http://dx.doi.org/10.1198/016214502388618960>
- Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405-430. <http://dx.doi.org/10.1002/for.928>
- Terui, N., & van Dijk, H.K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, 18, 421-438. [http://dx.doi.org/10.1016/S0169-2070\(01\)00120-0](http://dx.doi.org/10.1016/S0169-2070(01)00120-0)
- Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, Vol. 1, Ch. 4. Graham Elliot, Clive W.J. Granger, and Allan Timmermann eds. Elsevier B.V. [http://dx.doi.org/10.1016/S1574-0706\(05\)01004-9](http://dx.doi.org/10.1016/S1574-0706(05)01004-9)
- Trapletti, A. & Hornik, K. (2012). tseries: Time Series Analysis and Computational Finance. <http://CRAN.R-project.org/package=tseries>.
- Wei, X. & Yang, Y. (2012). Robust forecast combinations. *Journal of Econometrics*, 166, 224-236. <http://dx.doi.org/10.1016/j.jeconom.2011.09.035>
- Zellner, A., Hong, C. & Min, C.-k. (1991). Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics* 49, 275-304. [http://dx.doi.org/10.1016/0304-4076\(91\)90016-7](http://dx.doi.org/10.1016/0304-4076(91)90016-7)
- Zou, H. & Yang, Y. (2004). Combining time series models for forecasting, *International Journal of Forecasting*, 20, 69-84. [http://dx.doi.org/10.1016/s0169.2070\(03\)00004-9](http://dx.doi.org/10.1016/s0169.2070(03)00004-9)

Notes

Note 1. An institution from the Spanish Ministry of Inner Affairs that, among other things, is in charge of all aspects related to the road safety in Spain.

Note 2. We first estimated the models and obtained the forecasts to construct the panel data and form the panel decomposition. We carried out the estimation with SCATM by exact ML because, in this case, it is more efficient. However, using LS do not change our conclusions. Moreover, with a very large sample as ours, normality of the innovations is very plausible. In fact, tests applied to residuals do not reject the hypothesis.

Note 3. Issler and Lima (2009, section 2.1) consider a continuous set of models and split the total number of models J into M classes, each of them containing m nested models, so that $J = mM$. Hence, models within each class are nested, but models across classes are not. For each class of nested models, innovations can exhibit high cross-sectional dependence, but cross-sectional dependence of forecast errors across classes is expected to be weak. By choosing a large enough number of diversified classes of models, keeping some nested models poses no problem since the mixture of models will still deliver the optimal forecast. In small samples with few models, however, we should avoid nested models.

Note 4. The first subsample is included here because in this Application we did not have “directly available forecasts.” We also applied our panel decomposition to data provided by the Survey of Forecasters. In this case, however, there was no need to consider an estimation period because we already had at our disposal the forecasts.

Note 5. They are: Annual Percentage Error (APE), Forecasted Growth Rate, (FGR), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (sMAPE), and Symmetric Median Absolute Percentage Error (sMdAPE). See García-Ferrer et al. (2004) for the first two.

Note 6. All the forecasting results for individual models are also available from the authors upon request.

Note 7. Similar results were obtained in forecasting different quarterly data from the Survey of Professional Forecasters website. Separate results for the forecasts obtained for each year, forecast horizon, variable, model, combination, and accuracy measure are available from the authors upon request.

Note 8. Since July, 2007, and established very high traffic fines. This fact caused a break in the evolution of the variables.

Note 9. The “stationary” bootstrap is a block resampling with block lengths generated from a geometric distribution with a given mean length. Other alternatives are the “fixed” and “model-based” resampling. The former is a block resampling with “fixed” block length: Each replicate time series is found by taking blocks of fixed length from the original time series and putting them end-to-end until a new series of length given by the length of the original time series is created. The latter, which is very similar to the parametric bootstrap, avoids the complicated problem of choosing the block length, but relies on an accurate model choice being made.

Note 10. R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>. See also Davison & Hinkley (1997), Kim (2009), Canty & Ripley (2011), and Trapletti & Hornik (2012).

Note 11. Results for the other variables in this Application are available from the authors upon request.

Note 12. Bootstrapped biases and SE of weights for all models (6), months (12), forecast horizons (3), and variables (8) are available from the authors upon request. In general, the results are very similar to those mentioned here.