Centre for Growth & Business Cycle Research

Discussion Paper Series

Comparing Seasonal Forecasts of Industrial Production

By

Pedro M.D.C.B. Gouveia, Denise R. Osborn, Paulo M.M. Rodrigues

Centre for Growth and Business Cycle Research, Economic Studies,
University of Manchester, Manchester, M13 9PL, UK

May 2008
Number 102

COMPARING SEASONAL FORECASTS OF INDUSTRIAL PRODUCTION*

Pedro M.D.C.B. Gouveia
ESGHT
University of Algarve

Denise R. Osborn†
CGBCR
University of Manchester

Paulo M.M. Rodrigues
Faculty of Economics
University of Algarve

May 2008

Abstract

Forecast combination methodologies exploit complementary relations between different types of econometric models and often deliver more accurate forecasts than the individual models on which they are based. This paper examines forecasts of seasonally unadjusted monthly industrial production data for 17 countries and the Euro Area, comparing individual model forecasts and forecast combination methods in order to examine whether the latter are able to take advantage of the properties of different seasonal specifications. In addition to linear models (with deterministic seasonality and with nonstationary stochastic seasonality), more complex models that capture nonlinearity or seasonally varying coefficients (periodic models) are also examined. Although parsimonious periodic models perform well for some countries, forecast combinations provide the best overall performance at short horizons, implying that utilizing the characteristics captured by different models can contribute to improved forecast accuracy.

Keywords: Forecast accuracy, forecast combinations, seasonality, RMSPE, periodic models.

JEL classification: C22, C53.

---

*The first and third authors are grateful to Fundação para a Ciência e Técnologia for financial support through the POCTI program and FEDER (grant ref. no. POCTI/ECO/49266/2002). An earlier version of this paper was presented by the second author as a keynote address at the Conference on Seasonality, Seasonal Adjustment and their Implications for Short Term Analysis and Forecasting, organised by Eurostat in Luxembourg, May 2006. The helpful comments of conference participants are gratefully acknowledged.

†Corresponding author: Denise Osborn, Centre for Growth and Business Cycle Research, Economics, School of Social Sciences, University of Manchester, UK. Email: denise.osborn@manchester.ac.uk

1. Introduction

Agents working with seasonal data often require forecasts of intra-year observations; for example, managers need to forecast future monthly demand for their products in order to ensure that they have sufficient stocks on hand to meet this demand. Indeed, the production of many commodities is itself highly seasonal, largely due to the traditional factory closures that take place during the summer and Christmas periods. Perhaps because of their marked intra-year patterns, economists interested in seasonality have often focused on industrial production series (for example, Beaulieu and Miron, 1991, Cecchetti and Kashyap, 1996, Matas-Mir and Osborn, 2004).

The nature of seasonality is also of interest to official statistical agencies, including Eurostat, which is responsible for data provision relating to the European Union. Although many economists concentrate on seasonally adjusted values, the process of seasonal adjustment may itself involve forecasting future intra-year values of the unadjusted series, as discussed by Ghysels, Osborn and Rodrigues (2006) in the context of the recently-developed X-12-ARIMA method of the US Bureau of the Census. There has, however, been surprisingly little empirical analysis of the accuracy of methods for forecasting seasonal economic time series.

Rather than selecting a single model for forecasting, an alternative approach is to combine forecasts derived from a range of models. This has particular attraction in the context of forecasting seasonal series, since there are a number of different ways of handling seasonality that may be appropriate depending on the properties of the series in question. For example, seasonality may be of the deterministic form, it may exhibit nonstationary stochastic properties, it may be periodic (seasonally-varying coefficient) in nature, or it may exhibit non-linear interaction with the business cycle; see Ghysels and Osborn (2001) for discussion of some relevant models and their properties. Rather then choosing between these possibilities, a user may elect to adopt a forecast based on a combination of models. Indeed, the use of a suitably chosen combination may lead to improved forecast accuracy compared to the choice of a single method.

Since the early work of Bates and Granger (1969), several methods have been developed for combining forecasts. Since time series models are simplifications of complicated processes that are imperfectly understood, single models are typically incomplete representations of a data generating process (DGP). Hence, combinations of forecasts from different models, which may provide complementary information, can assist in the approximation of the DGP. In practice, such combinations are often found to outperform forecasts produced by a single model (see, *inter alia*, Bates and Granger, 1969, Granger and Ramanathan, 1984, Stock and Watson, 1999, 2004). Against this, Hibon and Evgeniou (2005) find that the best individual forecast model performs as well as the best combination. Nevertheless, as these authors state, combining forecasts retains an advantage in being less risky than selecting among the available individual model forecasts.

This paper studies the post-sample accuracy of forecasts of seasonally unadjusted monthly industrial production indices (IPI) from 17 individual countries (Austria, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Japan, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom and

USA) and an aggregate series for the Euro Area. In total, we examine 17 (linear and nonlinear) forecasting models and 18 procedures for combining the information from these 17 models. Our aim is to examine the relative accuracy of these approaches and to investigate whether any general lessons emerge about whether combining forecasts improves accuracy for these seasonal series.

The outline of the paper is as follows. Section 2 briefly introduces the forecast models considered in this paper. In Section 3 we study the empirical properties of the IPI series to investigate whether they display non-linearity and/or periodicity (seasonally-varying coefficients). Section 4 discusses predictive accuracy measures and introduces the combination methods considered. Substantive results in relation to forecast accuracy for the seasonal IPI series are contained in Section 5. Finally, Section 6 concludes the paper.

2. The Models

Our discussion first considers representations of seasonality in the context of constant parameter linear models, with subsequent subsections considering non-linear (SETAR) and periodic models. Although most of this discussion is general in the sense of referring to S seasons per year, our empirical analysis of monthly IPI below (obviously) employs $S = 12$.

2.1. Linear Seasonal Models

Following Ghysels, Osborn and Rodrigues (2006), for the purpose of presentation we write the seasonal model as

$$y_{Sn+s} = \mu_{Sn+s} + x_{Sn+s} \tag{2.1}$$

$$\phi(L)\, x_{Sn+s} = u_{Sn+s} \tag{2.2}$$

where y_{Sn+s} $(s = 1, ..., S, n = 0, ..., N-1)$ represents the observed value in season s (in our case a month) of year n, with total available observations assumed to be $T = SN$; the polynomial $\phi(L)$ contains any unit roots in y_{Sn+s} and is specified in the following subsections according to the model being discussed, L represents the conventional lag operator such that $L^k x_{Sn+s} \equiv x_{Sn+s-k}$, $k = 0, 1, ...,$ the driving shocks $\{u_{Sn+s}\}$ of (2.2) are assumed to follow an ARMA(p, q), $0 \leq p, q < \infty$ process written as $\beta(L) u_{Sn+s} = \theta(L) \varepsilon_{Sn+s}$, where the roots of $\beta(z) \equiv 1 - \sum_{j=1}^{p} \beta_j z^j = 0$ and $\theta(z) \equiv 1 - \sum_{v=1}^{q} \theta_v z^v = 0$ lie outside the unit circle, $|z| = 1$, and $\varepsilon_{Sn+s} \sim iid(0, \sigma^2)$. The term μ_{Sn+s} represents a deterministic kernel which is usually assumed to be either i) a set of seasonal means, *i.e.*, $\sum_{s=1}^{S} \delta_s D_{s,Sn+s}$ where $D_{s,Sn+s}$ is a dummy variable taking value 1 in season s and zero elsewhere, or ii) a set of seasonals with a (nonseasonal) time trend, *i.e.*, $\sum_{s=1}^{S} \delta_s D_{s,Sn+s} + \tau\,(Sn + s)$. In general, the second of these is more plausible for economic time series, since it allows the underlying level of the series to trend over time, whereas $\mu_{Sn+s} = \delta_s$ implies a constant underlying level, except for seasonal variation.

Linear forecasting models can be classified in terms of their assumptions about unit roots in $\phi(L)$. The two most common forms of seasonal models in empirical analyses employ either seasonally integrated models, with $\phi(L) = \Delta_S$ in (2.2), or deterministic seasonality combined with $\phi(L) = 1 - L$ or $\phi(L) = 1$. In addition, seasonal autoregressive integrated moving average (SARIMA) models with $\phi(L) = \Delta_S \Delta_1$ retain an important role as a forecasting benchmark. Each of these three models is briefly discussed in a separate subsection below.

2.1.1. Seasonally Integrated Model

The seasonally integrated model assumes that seasonality is nonstationary, with seasonal (or annual) differencing of y_{Sn+s} required in order to render the process stationary. Therefore $\phi(L) = 1 - L^S = \Delta_S$ in (2.1) and, since $\Delta_S = (1-L)(1 + L + L^2 + ... + L^{S-1})$, seasonal integration imposes the presence of unit roots not only at the zero frequency, but also at each of the so-called seasonal frequencies.

Stationary dynamics in economic time series are often represented in autoregressive (AR) form. With this assumption, namely $\beta(L)u_{Sn+s} = \varepsilon_{Sn+s}$ in (2.2) with $\beta(L)$ a pth order polynomial, the seasonally integrated model is

$$\beta(L)\Delta_S y_{Sn+s} = \beta(1)S\tau + \varepsilon_{Sn+s} \tag{2.3}$$

since $\Delta_S \mu_{Sn+s} = S\tau$. Thus, with the inclusion of an intercept, the seasonally integrated process of (2.3) has a common annual drift, $\beta(1)S\tau$, across seasons. Clearly, the essential features of the model are retained if a moving average component is added to (2.3). Notice that the underlying seasonal means μ_{Sn+s} are not observed, since the seasonally varying component $\sum_{s=1}^{S} \delta_s D_{s,Sn+s}$ of μ_{Sn+s} in (2.1) is annihilated by seasonal (that is, annual) differencing.

From an economic point of view, nonstationary seasonality can be controversial because the values over different seasons are not cointegrated and hence can move in any direction in relation to each other, so that "*winter can become summer*". This lack of cointegration appears to have been first noted by Osborn (1993).

2.1.2. Deterministic Seasonal Models

When seasonality results in peaks and troughs across particular seasons year after year, it may be described by deterministic variables leading to what is conventionally referred to as *deterministic seasonality*. In this case the underlying seasonal pattern is assumed to display means that are constant over time.

A simple deterministic seasonal model with stationary AR dynamics can be given as

$$\beta(L)y_{Sn+s} = \sum_{s=1}^{S} \beta(L)\left(\delta_s D_{s,Sn+s}\right) + \beta(1)(Sn+s)\tau + \varepsilon_{Sn+s} \tag{2.4}$$

where ε_{Sn+s} is again a zero mean white noise process. In this case, the deterministic component of the estimated model explicitly includes seasonal intercepts and a linear trend. However, the assumption of stationary dynamics may be unrealistic since it is common for economic time series to exhibit evidence of a zero frequency unit root. Therefore, $\phi(L) = 1 - L$ may be imposed in (2.2). Again assuming that the stationary dynamics are of the AR form, (2.1)-(2.2) then becomes

$$\beta(L)\Delta_1 y_{Sn+s} = \sum_{s=1}^{S} \beta(L)\Delta_1 \mu_{Sn+s} + \varepsilon_{Sn+s} \tag{2.5}$$

where $\Delta_1 \mu_{Sn+s} = \mu_{Sn+s} - \mu_{Sn+s-1}$, so that (only) the change in the seasonal means between seasons s and $s-1$ is identified.

2.1.3. SARIMA Model

When working with nonstationary seasonal data, both annual changes and changes between adjacent seasons are important concepts. This motivates the model

$$\beta(L)(1-L)(1-L^S)y_{Sn+s} = \theta(L)\varepsilon_{Sn+s} \tag{2.6}$$

which results from specifying $\phi(L) = \Delta_1\Delta_S = (1-L)(1-L^S)$ in (2.2). The intuition is that the filter $(1-L^S)$ captures the tendency for the value of the series for a particular season to be highly correlated with the value for the same season a year earlier, while $(1-L)$ captures the nonstationary nonseasonal stochastic component. It is worth noting that the imposition of $\Delta_1\Delta_S$ annihilates the deterministic variables (seasonal means and time trend) of (2.1), so that these do not appear in (2.6).

However, since $(1-L)(1-L^S) = (1-L)^2(1+L+L^2+...+L^{S-1})$, (2.6) imposes unit roots at all seasonal frequencies, as well as two unit roots at the zero frequency. As a result these models may be empirically implausible (see *e.g.* Osborn, 1990, and Hylleberg, Jørgensen and Sørensen, 1993). Nevertheless, they can be successful in forecasting due to their parsimonious nature and hence provide a useful benchmark for forecast accuracy comparisons.

A specific SARIMA model of particular interest as a benchmark for seasonal forecasting is the "airline model" of Box and Jenkins (1970), which imposes $\beta(L) = 1$ in (2.6), together with $\theta(L) = (1-\theta L)(1-\Theta L^S)$.

2.2. Seasonal SETAR Models

The SETAR class of non-linear models allows the classification of observations into different regimes according to the value taken by a specific threshold variable, and hence captures a form of asymmetric or time-varying behaviour.

In this study, we consider a seasonal two-regime SETAR ($SSETAR$) model of order p of the form

$$y_{Sn+s} = \sum_{k=1}^{2}\sum_{s=1}^{S}[\delta_{s,k}D_{s,Sn+s} + \tau_{s,k}D_{s,Sn+s}(Sn+s)]I_{k,Sn+s} + \sum_{i=1}^{p}\rho_i y_{Sn+s-i} + \varepsilon_t \tag{2.7}$$

where $I_{k,Sn+s}$, $k = 1, 2$, corresponds to a binary indicator variable with value determined by the threshold variable q_{Sn+s-d}; the disturbance in each regime is assumed to be white noise with constant variance. The regimes in (2.7) are defined by the value of q_{Sn+s-d} in relation to some constant threshold γ. In practice the threshold variable is typically a lag of y_{Sn+s}, or a linear combination of lagged

y_{Sn+s}; see, *inter alia*, Tsay (1989, p.23) and Hansen (1997, p.10). Notice that (2.7) allows the intercept and trend to vary with the regime as well as the season, with the AR lag coefficients assumed to be time-invariant. Although more general forms of SSETAR model can be employed, their greater flexibility implies the estimation of a larger number of coefficients. In our forecasting context, we prefer the more parsiminous version of (2.7).

Implicitly (2.7) assumes that y_{Sn+s} is stationary. To allow for unit root behavior, (2.7) may be estimated using $\Delta_1 y_{Sn+s}$ (as in Matas-Mir and Osborn, 2004) or $\Delta_S y_{Sn+s}$, with appropriate changes in the deterministic terms of the regression.

However, before this type of model is applied in empirical work, it is important to determine whether the data justifies its use through a test for threshold effects. Chan (1990) and Hansen (1997) suggest the test statistic

$$F(\gamma) = T\left(\frac{\widetilde{\sigma}^2 - \widehat{\sigma}^2(\gamma)}{\widehat{\sigma}^2(\gamma)}\right) \tag{2.8}$$

where $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2(\gamma)$ represent the disturbance variance estimators acquired from the residuals of a linear and a SETAR model, respectively. The null hypothesis considers a linear model as appropriate, while the alternative of regime-dependent coefficients supports the SETAR model. A difficulty in applying these tests arises when γ is unknown since this parameter is identified only under the alternative hypothesis. This problem was first identified by Davis (1977); see also Hansen (1996). As the asymptotic distributions are non-standard, critical values for (2.8) in a specific application can be obtained using the bootstrap method suggested by Hansen (1997).

2.3. Periodic Models

Periodic autoregressive (PAR) models provide a class of model for seasonally unadjusted data which allow the coefficients to change according to the seasons of a year. This seasonal parameter variation can prove useful in describing economic situations in which choices made by economic agents show distinct seasonal characteristics. Problems associated with dismissing periodicity are well described in Osborn (1991) and in Tiao and Grupe (1980).

PAR models assume that the observations for different seasons can be described by distinct autoregressive models. We consider the following PAR(S,p) model

$$y_{Sn+s} = \sum_{s=1}^{S}[\delta_s D_{s,Sn+s} + \tau_s D_{s,Sn+s}(Sn+s)] + \sum_{s=1}^{S}\sum_{j=1}^{p_s}\alpha_{js} D_{s,Sn+s} y_{Sn+s-j} + \varepsilon_{Sn+s} \tag{2.9}$$

where p_s is the order of the autoregressive component corresponding to season s, $p = \max(p_1, ..., p_S)$, and $\varepsilon_{Sn+s} \sim iid\,(0, \sigma^2)$. In its unrestricted form of (2.9), the model coefficients can be estimated by ordinary least squares. PAR models can be applied to either the levels of the series, as in (2.9), or after the application of first differences. Also, the trend coefficient can either be allowed to vary with the season or to be constant with $\tau_s = \tau$ $(s = 1, ..., S)$.

Similarly to the SETAR models previously discussed, it is advisable also in this case to verify whether the data shows this type of property before employing a PAR forecasting model. Allowing the possibility of a seasonal deterministic component, the most direct test of the non-periodic null hypothesis considers

$$H_0 : \alpha_{is} = \alpha_i, \quad s = 1, ..., S, i = 1, ..., p \tag{2.10}$$

against the alternative that not all α_{is} are equal, which we denote as F_{PAR}. This test can be performed by the usual F-test, which (for T sample observations used for estimation of (2.9)) asymptotically follows an F distribution with $((S-1)\,p,$ $T-(S+2)p)$ degrees of freedom; see Boswijk and Franses (1996). Alternatively, following Franses (1996, pp.101-102) a residual-based approach can be adopted. As a first step, a non-periodic AR(p) model is estimated for $\Delta_S y_{S+s}$. Using the resulting residuals, periodicity is tested through the auxiliary regression,

$$\widehat{v}_{Sn+s} = \sum_{i=1}^{p} \phi_i \Delta_S y_{Sn+s-i} + \sum_{j=1}^{r} \sum_{s=1}^{S} \gamma_{js} D_{s,Sn+s} \widehat{v}_{Sn+s-j} + u_{Sn+s} \tag{2.11}$$

via an F-test for the joint significance of the γ_{js} for order r. Under the non-periodic null hypothesis, this F-statistic asymptotically follows a standard F-distribution with $(Sr, T-p-Sr)$ degrees of freedom.

As an additional option, or a complementary procedure, the auxiliary regression

$$\widehat{v}^2_{Sn+s} = \omega_0 + \sum_{k=1}^{S-1} \omega_k D_{k,Sn+s} + e_{Sn+s} \tag{2.12}$$

can be used to check for seasonal heteroscedasticity. As argued by Franses (1996), neglected parameter variation may surface in the variance of the residual process. Under the null hypothesis of no seasonal heteroscedasticity, an F-test for $\omega_k = 0, k = 1, ..., S-1$ asymptotically follows a standard F-distribution with $(S-1, T-p)$ degrees of freedom. It should be noted, however, that finding of seasonal heteroscedasticity does not necessarily imply that a PAR model should be used, since this could arise from a conventional constant-coefficient model subject to disturbances which have seasonally-varying variances.

3. Empirical Properties of Industrial Production

3.1. Data

The data used in this study is the logarithm of monthly IPI data for 17 individual countries and the Euro Area. Therefore, the first difference has the interpretation of the monthly growth rate, and the annual difference as the annual growth rate. Table 1 reports some descriptive statistics for these annual and monthly growth rates, after outlier correction. Outlier detection and correction was carried out using the Tramo/Seats program developed by Gømez and Maravall (1996). For ease of interpretation, the differenced values are multiplied by 100 prior to the calculation of the statistics of Table 1.

Table 1: Descriptive Statistics

Country	Outliers			Annual Growth		Monthly Growth	
	AO	TC	LS	Mean	SD	Mean	SD
Austria	0	0	0	3.21	3.90	0.30	8.59
Canada	0	1	0	2.57	4.87	0.19	6.25
Denmark	0	2	0	2.40	6.55	0.20	17.16
Euro Area	3	0	3	1.55	3.09	0.12	11.12
Finland	3	0	2	3.39	5.81	0.29	14.92
France	0	0	1	1.14	3.15	0.08	15.13
Germany	2	0	1	1.37	3.37	0.10	6.71
Greece	2	3	1	1.09	4.11	0.09	8.04
Hungary	0	1	0	2.52	8.72	0.25	10.90
Italy	0	0	0	0.90	4.36	0.06	31.85
Japan	0	0	0	1.68	4.91	0.18	7.53
Luxembourg	1	1	2	3.29	6.01	0.24	12.53
Netherlands	0	0	0	0.99	4.16	0.07	7.84
Portugal	1	0	0	2.44	4.99	0.18	15.59
Spain	0	0	1	1.70	3.46	0.17	21.00
Sweden	3	0	2	2.64	4.52	0.22	26.46
UK	0	0	1	1.00	3.52	0.06	7.14
USA	2	2	1	2.52	3.67	0.21	2.01

Note: The columns labeled AO, TC and LS refer to the nature of outliers detected and indicate the respective number of outliers observed. Outlier detection and correction was carried out using the automatic procedure in TRAMO/SEATS.

Our data covers the period January 1980 to December 2005 (before differencing). However, outliers are removed only for the subsample used for the estimation of the models, which is January 1980 to December 2002.

Although IPI growth over this period is positive in all cases, Table 1 indicates very different experiences across the countries considered for its mean and variability. Indeed, Italy, Spain and Sweden have a standard deviation of monthly growth around six to eight times that of annual growth, pointing to the highly seasonal nature of these IPI series. On the other hand, for Canada, Hungary and USA, these monthly and annual growth rate standard deviations are of similar magnitude.

The remainder of this section discusses some tests undertaken to examine the characteristics of our data series. The outlier corrected subsample to December 2002 is used for this analysis.

3.2. Nonlinearity

As an indicator of the potential value of SETAR models for our seasonal series, we test for the presence of threshold effects in the data by investigating whether the difference between the coefficients in the regimes is significant in (2.7).

Order selection for the autoregressive component of the seasonal SETAR model is also important. In this study, the order of the test regressions employed was determined following a general-to-specific procedure (see Ng and Perron, 1995) in a linear AR model, using a maximum lag of $p = 12$. The p-values for the linearity test are obtained using the bootstrap, as suggested by Hansen (1997) in this context. Our application uses 5000 bootstrap replications.

Table 2 presents the results of tests of linearity against SETAR type nonlinearity, which use a regression of the form of (2.7) with $q_{12n+s-d} = \Delta_{12}y_{12n+s-d}$. Both the delay d and the threshold value γ are endogenously determined, with the latter obtained from a search over the central 70% of the empirical distribution of $\Delta_{12}y_{12n+s-d}$. The selected AR order (given in the column *AR order* in Table 2) is also the maximum delay permitted for d.

Table 2: Testing for Linearity in Industrial Production Series

Country	γ	d	Linearity test (p-value)	AR order
Austria	-.005	4	.016	12
Canada	-.015	1	.121	10
Denmark	.052	4	.052	12
Euro Area	.014	10	.001	12
Finland	.030	12	.218	12
France	.019	8	.167	12
Germany	.037	5	.007	12
Greece	.027	5	.037	12
Hungary	.117	9	.023	12
Italy	.078	12	.112	12
Japan	-.026	2	.054	10
Luxembourg	.094	8	.024	12
Netherlands	-.065	6	.029	12
Portugal	-.003	7	.070	12
Spain	.014	6	.464	12
Sweden	.021	9	.001	12
UK	.045	1	.129	10
USA	-.051	2	.067	9

From Table 2, we observe that (at a 5% level) the linear structure is rejected for nearly half of the countries considered, with the strongest evidence of nonlinearity being for Germany, Sweden and the Euro Area. These results may be indicative of the presence of an interaction between seasonality and the business cycle, as discussed by Matas-Mir and Osborn (2004) for industrial production series.

3.3. Periodicity

Given that monthly data are used in this paper, an obvious PAR model to consider is a $PAR(12, p)$, which identifies each month as a distinct "season". However, if used without restrictions, this model has the potential disadvantage of being highly parameterised. For instance, an unrestricted $PAR(12, 12)$, even with no deterministic terms, requires estimation of $12 \times 12 = 144$ coefficients, which points to identifying more parsimonious models.

One strategy adopted below is based on the assumption that common behaviour is present for specific months, leading to the proposal of a $PAR(3,p)$ model. These $PAR(3,p)$ models are determined by defining three distinct groups of months, with one group including all months with negative average monthly growth, another all months with relatively low monthly positive growth and the third group includes months with the strongest observed average monthly growth[1]. This grouping leads to three "seasons" with different numbers of observations.

Table 3 presents the results of tests for periodic coefficient variation. The test procedures denoted as F_{PeAR1_12} and F_{SH} are applied to the residuals of an AR model as described in (2.11) and (2.12) fitted to the annually differenced series. The F_{PAR} test examines the non-periodic null hypothesis of (2.10) in a regression for the levels of the data. The maximum lag length p considered is 12 in both cases, with this order reduced where appropriate through a testing down strategy.

Table 3: Testing for Periodicity in Industrial Production Series

	$FPeAR1\_12$		F_{SH}		F_{PAR}	
	PAR(3,p)	PAR(12,p)	PAR(3,p)	PAR(12,p)	PAR(3,p)	PAR(12,p)
Austria	1.83*	1.37*	3.69*	3.50*	2.05*	1.52*
Canada	2.73*	1.55*	0.08	1.5	8.24*	3.07*
Denmark	3.35*	2.79*	3.14*	1.66	0.64	1.92*
Euro Area	2.81*	1.99*	1.72	1.33	0.88	4.26*
Finland	1.59*	1.66*	3.08	1.79	1.14	4.51*
France	1.85*	2.00*	1.84	1.42	4.94*	3.15*
Germany	2.64*	1.89*	2.6	1.59	1.71	2.22*
Greece	3.50*	1.88*	1.82	1.63	3.69*	1.89*
Hungry	2.46*	2.02*	0.46	0.36	1.02	2.87*
Italy	1.84*	1.91*	3.69*	6.80*	1.06	2.62*
Japan	3.80*	1.88*	0.17	1.59	3.73*	2.60*
Luxemburg	1.80*	1.75*	2.41	1.35	1.55	1.76*
Netherlands	2.84*	2.84*	0.32	0.97	1.1	2.95*
Portugal	2.45*	1.56*	2.53	1.31	1.56	2.10*
Spain	2.23*	2.05*	0.48	0.97	4.61*	4.44*
Sweden	3.15*	2.24*	3.57*	2.73*	3.02	5.82*
UK	2.46*	2.35*	0.92	1.72	1.41	3.21*
USA	1.76*	1.74*	1.12	1.21	7.37*	1.30*

Note: * indicates significance at the 5% level.

Whether applied to the residuals or the levels of the data, Table 3 provides strong evidence in favour of the presence of periodic coefficient variation across the IPI series analysed. When directly applied to the coefficients, however, the F_{PAR} test rejects substantially less often when data are grouped in three seasons than when all 12 months are considered separately. Nevertheless, the regression

[1] To be specific, the positive growth regime classifications generally used an average growth of less than 0.1% per month as low growth and monthly growth of 0.1% or more as high growth. However, different classification rules were used for Canada and the US, for which low (positive) growth was defined in relation to thresholds of 0.03% and 0.05%, respectively.

on which the latter test is based is highly parameterised, so this result may not be entirely reliable. The F_{SH} test provides only weak evidence of seasonal heterocedasticity across the series considered. Nevertheless, the other results point to the potential value of using periodic models for forecasting.

4. Forecast Accuracy

To study forecast accuracy, we consider the use of m post-sample observations to evaluate h-step ahead forecasts generated from models fitted to the first T observations.

Although many measures of forecast accuracy are available, we follow much of the literature in basing our evaluation on the Root Mean Squared Prediction Error ($RMSPE$), defined as

$$RMSPE(h) = \sqrt{\frac{1}{m-h+1}\sum_{j=h}^{T+m}\left(\widehat{y}_{T+j|T+j-h} - y_{T+j}\right)^2} \tag{4.1}$$

where $\widehat{y}_{T+j|T+j-h}$ is the h-step ahead forecast made for period $T+j$ based on data available at $T+j-h$. In order to focus on the role of seasonality, which may be anticipated to be most marked for short-term forecasts of less than a year, results are computed for horizons $h = 1, 3, 8$.

To reflect the different approaches to seasonality, we consider specifications based on various levels of differencing, namely no differences, together with models after application of the filters Δ_S, Δ_1 and $\Delta_S\Delta_1$. All forecasts are considered in relation to the level of the IPI series using RMSPE as defined in (4.1).

There is growing empirical evidence that nonlinear models perform useful information for forecasting (see, *inter alia*, Terui and van Dijk, 2002). Clements *et al.* (2003) compare linear autoregressive and SETAR models and study the degree of non-linearity that needs to be present in the data before forecasts from non-linear models outperform linear rivals. Although there are relatively few studies of the forecast performance of periodic models, evidence provided by Osborn and Smith (1989) and Rodrigues and Gouveia (2004) indicates that they can perform well if appropriately specified.

4.1. Nonlinear Model Forecasts

For the nonlinear (seasonal) SETAR models of (2.7), computing point forcasts is considerably more involved than computing forcasts from linear models. To illustrate this, consider the case where the variable y_{Sn+s} is described by a first order nonlinear autoregressive model which is summarised as

$$y_{Sn+s} = F\left(y_{Sn+s-1}; \theta\right) + \varepsilon_{Sn+s}. \tag{4.2}$$

In this context, when the forecast horizon is longer than 1 period, the linear conditional expectation operator can not be interchanged with the nonlinear operator F, since $E\left[F\left(.\right)\right] \neq F\left(E\left[.\right]\right)$. Consequently, for a given parameter vector θ

and horizon h

$$E\left[F\left(y_{Sn+s+h-1};\theta\right)|\Omega_{Sn+s}\right] \neq F\left(E\left[\left(y_{Sn+s+h-1|Sn+s}\right)|\Omega_{Sn+s}\right];\theta\right), \qquad h>1$$

where Ω_{Sn+s} indicates information available at time $Sn+s$. Obtaining an unbiased point forecast based on (4.2) requires estimation of the left-hand side of this expression.

The distribution of the white noise disturbance ε_{Sn+s} in (4.2) is never known with certainty. However, to overcome this difficulty, forecasts can be computed using Monte Carlo or bootstrap methods. Clements and Smith (1997) compare various methods to obtain multiple-step-ahead forecasts for SETAR models and conclude that the bootstrap method compares favourably to the other methods. This method is also adopted here, where we use 500 bootstrap replications in order to approximate $E\left[F\left(y_{Sn+s+h-1};\theta\right)|\Omega_{Sn+s}\right]$.

4.2. Methods for Forecast Combinations

Combining forecasts, as introduced by Bates and Granger (1969), has often been found to improve forecast accuracy compared with using an individual forecasting method. The effectiveness of simple averaging is demonstrated by, among others, Bates and Granger (1969) and Granger and Ramanathan (1984), while other studies demonstrate the usefulness of other approaches to combining multiple individual forecasts; see, *inter alia*, Armstrong (1989); Clemen (1989), Diebold and Lopez (1996), Hendry and Clements (2002), Newbold and Harvey (2002), Stock and Watson (1999) and Terui and van Dijk (2002).

The remainder of this section briefly introduces the forecast combination methods that we employ in our empirical analysis. In addition to simple averages (mean or median), methods for combining forecasts can be classified as being based on historical RMSPE or derived from regression methods. We devote separate subsections to each of these approaches.

4.2.1. Historical RMSPE

The combinations based on historical RMSPE are formed as a weighted average of individual forecasts, with weights varying with the historical performance of each individual forecast; see, *inter alia*, Diebold and Pauly (1987) and Stock and Watson (1999). For k separate h-step forecasts, namely $\widehat{y}^{i}_{Sn+s+h|Sn+s}$ $(i=1,...,k)$, the forecast combination is given by

$$\widehat{y}^{c}_{Sn+s+h|Sn+s}=\sum_{i=1}^{k} w_i^h \widehat{y}^{i}_{Sn+s+h|Sn+s}$$

where the weight w_i^h is

$$w_i^h=\frac{RMSPE(h)_i^{-\lambda}}{\sum\limits_{j=1}^{k} RMSPE(h)_i^{-\lambda}} \tag{4.3}$$

and $RMSPE(h)_i$ is the Root Mean Square Predictor Error for method i at horizon h. Since the relative performance of different models can change over time, following Bates and Granger (1969) we compute RMSPE at the end of the sample T using information relating to forecasts for the final three years of the estimation sample, namely $T-35$ to T.

As implied by (4.3), the weights on the constituent forecasts are inversely related to their RMSPE values. A simple average which places equal weight on all forecasts corresponds to $\lambda = 0$. As λ increases, more weight is placed on those models that have been performing relatively well. In this paper we consider $\lambda \in \{0, 1, 1.25, 1.5, 2\}$, with $\lambda = 2$ implying that the weights are inversely proportional to the mean square prediction error.

In addition to weighting as in (4.3), we also consider weights that discount historical forecast accuracy based on RMSPE (see Stock and Watson, 2004). For an h-step ahead forecast, the combination weight in this case has the form

$$w_i^h = \frac{m_i^{-h}}{\sum\limits_{j=1}^{k} m_j^{-h}} \tag{4.4}$$

where

$$m_i^h = \sqrt{\sum_{j=1}^{T-h} \delta^{j-1} \left(\widehat{y}^i_{T-j+1|T-j+1-h} - y_{T-j+1}\right)^2} \tag{4.5}$$

and δ is the discount factor. In this paper, we consider discount factors of $\delta \in \{1, 0.95, 0.90\}$. Note that for $\delta = 1$ (4.5) operates as a window with no discounting and this weighting scheme is then equivalent to (4.3) with $\lambda = 2$ when the latter uses all observations to time T.

4.2.2. Regression Methods

Granger and Ramanathan (1984), Diebold (1988) and others, suggest combining forecasts using regression methods. Following Diebold (2001, p.297) we do not force the weights to sum to unity, nor do we exclude an intercept. Indeed, inclusion of an intercept facilitates bias correction and allows biased forecasts to be combined. Therefore, the regression used can be expressed as

$$y_{Sn+s+h} = \beta_0 + \sum_{j=1}^{k} \beta_j \widehat{y}^j_{Sn+s+h|Sn+s} + \varepsilon_{Sn+s+h} \tag{4.6}$$

with $j = 1, 2, ..., k$, and where k represents the number of individual forecast methods included. The weights in (4.6) for the observation at T are estimated using the final 36 observations in the estimation period, namely corresponding to h-step ahead forecasts for periods $T-35$ to T inclusive.

However, with $k = 17$ in our case, (4.6) implies an excessive parameterisation. Therefore, in implementing (4.6) we use only the five methods producing the most accurate forecasts over the latest available 36 observations.

5. Forecast Performance

5.1. Individual Forecasting Models

Table 4 presents the individual forecasting models that we apply, while the 18 forecast combination procedures used are summarised in Table 5.

As noted in Section 2, the time series models are generally applied to macroeconomic data after differencing, and this is replicated in Table 4. Since the appropriate level of differencing is often unclear in empirical analyses, we reflect this uncertainty by applying the same models to data after applying each of the common differencing filters. For example, low-order ARMA models are applied to data after annual differencing, and to data after both first and annual differences are applied; PAR models are estimated for levels and first differenced data, while SETAR models employ levels and data after (separate) application of the Δ_1 and Δ_S filters. In each case these choices reflect the type of data to which these models are applied in practice, in conjunction with the indicated deterministic terms. For the $AR(p)$ models of Table 4, a maximum order of 24 is considered, while the $SETAR(p)$ considers a maximum AR order of 12. In both cases, insignificant lags are eliminated (starting with the minimum $t-statistic$) prior to using the models for forecasting. Based on a preliminary investigation, the lag order p is set to 3 for all PAR models.

Table 4: Forecast Models

Code	Individual Models	Filter	Deterministic terms
M1	$Airline\ Model$	$\Delta_1\Delta_{12}$	None
M2	$ARMA(1,1)$	Δ_{12}	Intercept
M3	$ARMA(2,2)$	Δ_{12}	Intercept
M4	$AR(p)$	Δ_{12}	Intercept
M5	$SSETAR\ (p)$	Δ_{12}	Intercept
M6	$AR\ (p)$	$levels$	Seasonal intercepts + trend
M7	$PAR(12,3)$	$levels$	Seasonal intercepts + trend
M8	$PAR(3,3)$	$levels$	Seasonal intercepts & seasonal trends
M9	$SSETAR(p)$	$levels$	Seasonal intercepts & seasonal trends
M10	$AR\ (p)$	Δ_1	Seasonal intercepts
M11	$PAR(3,3)$	Δ_1	Seasonal intercepts + trend
M12	$PAR(12,3)$	Δ_1	Seasonal intercepts & seasonal trends
M13	$SSETAR(p)$	Δ_1	Seasonal intercepts
M14	$ARMA(1,1)$	$\Delta_1\Delta_{12}$	None
M15	$ARMA(2,2)$	$\Delta_1\Delta_{12}$	None
M16	$ARMA(3,3)$	$\Delta_1\Delta_{12}$	None
M17	$AR(p)$	$\Delta_1\Delta_{12}$	None

5.2. Forecasting Combination Methods

The combination methods that we consider are based i) on the weight function (4.3) with $S = 12$ and $\lambda \in (0, 1, 1.25, 1.5, 2)$; ii) on the mean and median of the best 5, 10 and 15 models which are chosen based on the historical RMSPE across the 17 models considered; *iii*) on the discounted RMSPE weights as given in (4.4) and (4.5) with $S = 12$ and $\delta \in \{1, 0.95, 0.90\}$; *iv*) based on the regression method as described in (4.6) and; *v*) finally, on the mean or median of all combinations. Table 5 summarizes the individual combination methods used in the empirical analysis.

Table 5: Combination Methods

Code	Combination method	Parameters
C1	mean of M1 to M17	
C2	median of M1 to M17	
C3	(4.3)	$\lambda = 0$
C4	(4.3)	$\lambda = 1$
C5	(4.3)	$\lambda = 1.25$
C6	(4.3)	$\lambda = 1.5$
C7	(4.3)	$\lambda = 2$
C8	mean of best 5 models (RMSPE criteria)	
C9	mean of best 10 models (RMSPE criteria)	
C10	mean of best 15 models (RMSPE criteria)	
C11	median of best 5 models (RMSPE criteria)	
C12	median of best 10 models (RMSPE criteria)	
C13	median of best 15 models (RMSPE criteria)	
C14	mean of combinations	
C15	(4.4) and (4.5)	$\delta = 1$
C16	(4.4) and (4.5)	$\delta = 0.95$
C17	(4.4) and (4.5)	$\delta = 0.9$
C18	regression method of (4.6)	

5.3. Forecasting Results

Forecast accuracy is evaluated by employing the 36 observations from January 2003 to December 2005. All forecasting models are recursively re-estimated over this forecasting period. In addition, models that require specification of the appropriate AR order are recursively re-specified during the forecast period, while the weights required in (4.3), (4.4) and (4.6) are also updated using the most recently available observations and the corresponding forecast values.

5.3.1. Individual Countries

Table 6 presents the ten best forecasting approaches, according to post-sample RMSPE(h), out of the 35 considered for each of the countries examined, including the Euro Area. Not surprisingly, the best performing methods differ over both the horizon ($h = 1, 3, 8$) considered and the country. However, two cases of particular interest may be the USA and the Euro Area. For the former, the simple $AR(p)$ model M6 estimated in levels, with seasonal intercepts and a trend, does well at short horizons ($h = 1$, 3), but this model does not enter the best ten for $h = 8$. On the other hand, the PAR(3,3) model M11 in first differences is the most accurate method at a horizon of $h = 8$ months and provides a relatively good performance at $h = 3$, but does not enter the top ten at one month ahead. Overall, the best results for the USA are given by combining forecasts using the either the median or mean forecast from the most accurate 5 models (C11 and C8 respectively). In particular, this median combination yields the most accurate forecasts at horizons of one, a close competitor to the best at $h = 3$ and the fifth most accurate at $h = 8$.

[Insert Table 6 about here]

For the Euro Area, a good performance of the PAR(3,3) model M11 at $h = 1, 3$ and 8 is again observed. The closest competitor at very short horizons is another PAR model in levels, namely M7. For the Euro Area, however, it is striking that combinations perform well at $h = 8$, where a variety of combined forecasts are more accurate than any individual model except for the PAR(3,3).

Across all the sub-tables in Table 6, some interesting patterns emerge. Of the 64 evaluations (namely, 18 countries at three forecast horizons), individual models provide the most accurate forecasts (with a ranking of 1) in approximately two thirds of cases, with the PAR(3,3) model in first differences M11 achieving this most often, namely nine times. The next best performing individual model is the airline model M1, being top ranked six times, while a simple AR(p) in seasonal and annual differences, M17, also provides robust forecasts with five top places. Of the forecast combinations, the regression method C18 is best, being ranked top five times.

The relatively good performance of the PAR models reflects the evidence for such effects uncovered in Table 3. Indeed, these perform particularly well at the short forecast horizons of one and three months when estimated on the levels data, with either M11 or M12 the best forecast model at one or both of these horizons for Austria, Germany, Japan, Netherlands and Sweden, in addition to the Euro Area. On the other hand, however, the nonlinear SETAR models do not generally perform well according to the results of Table 6, although the variants M5 and M9 do yield the most accurate forecasts at $h = 1$ for Portugal and Luxembourg respectively. Nevertheless, overall, the evidence of nonlinearity in Table 2 does not here lead to improved post-sample forecast accuracy.

Despite individual models typically delivering the most accurate forecasts in Table 6, the implications change if the top ten models are considered. For $h = 1$, combinations occupy 82% of the top ten positions compared with 18% from the individual models; for $h = 3$, forecast combinations take 71% of the top ten positions; and for $h = 8$, the distribution becomes slightly more symmetric in that combinations represent 66% of these positions. Indeed, for a few countries (specifically, Denmark, Hungary and Spain), combinations occupy all of the top ten places for one-month ahead forecast accuracy, while for others (France, Germany, Japan, Luxembourg and the Netherlands) a specific model delivers the most accurate forecast at this horizon but combinations occupy all the remaining top ten places. These results imply that while some individual methods can provide good forecasts, combinations may deliver more consistent forecast performance.

5.3.2. Overall Accuracy Comparisons

To investigate the overall accuracy of these methods, Table 7 provides summary information on average rankings and RMSPE. The RMSPE results in the final column are scaled by that of M1 (the airline model) at each horizon in order to facilitate comparisons.

[Insert Table 7 about here]

Here some individual linear models perform reasonably well, including the airline model (M1) and M17, as well as the PAR models M7, M11 and M12. However, it should be noted that the airline and the AR models, M1 and M17, take account of seasonality only in the naive fashion of largely removing it through the application of annual (in addition to first) differences. Although a number of individual models provide more accurate forecasts than the airline model in terms of average relative RMSPE at $h = 1$, the robust performance of this model becomes notable at longer horizons in Table 7, with no individual model providing more accurate average relative RMSPE at $h = 8$. At shorter horizons, however, the PAR models, and especially the parsiminous PAR(3,3) model in first differences, M11, capture the seasonality more adequately than any other specification that explicitly models these patterns.

The poor performance of the $SETAR$ model in levels (M9) and first differences (M13) is particularly marked in Table 7. This may be a consequence of the nonlinearity often evident in Table 3 not being repeated during the forecast period. The results are also not favourable overall to the use of low order $ARMA$ specifications after first and annual differences (M14, M15, M16) for forecasting these IPI series.

The average rankings in Table 7 are, however, clear about the quality of the forecast combinations, with the overall average rankings for these being generally superior to the individual models considered. Indeed, combinations always occupy the top average ranking positions. For $h = 1$, the best performance in terms of relative RMSPE is given by the combinations based explicitly on past RMSPE, namely C16 and C17 (discounted RMSPE forecast weights based on (4.4) and (4.5) with weights $\delta = 0.95$ and $\delta = 0.90$ respectively), followed by C4, C5, C6 and C7 (all based on (4.3) for different choices of λ). In terms of average rank, however, C17 performs best at this horizon. For $h = 3$ the best overall average performance in terms of both average rank and relative RMSPE are given by C15 (undiscounted RMSPE forecasts based on (4.4) and (4.5) with $\delta = 1$) and C4 (weights (4.3) with $\lambda = 1$). For $h = 8$, the best performance is C16, followed by C4, C10 and C17. Overall, therefore, these results point to the use of forecast combination methods that use previous RMSPE performance as weights.

Table 7 also indicates that almost all combination methods considered give average RMSPE gains compared to even the best of the individual models for the short horizons $h = 1, 3$. The only exceptions to this are C1 (mean of all models) and C3 ((4.3) with $\lambda = 0$) and C18 (regression method of (4.6)).

6. Conclusion

This study reinforces evidence that combining forecasts from individual models can improve post-sample forcast accuracy. Our conclusion is based on forecasts from a large set of individual models and methods of combination of forecasts, whose performance is evaluated using monthly seasonally unadjusted Industrial Production data from 17 individual countries (Austria, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Japan, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom and USA) and an aggregate series for the Euro Area.

The potential of forecast combinations is even more attractive in the context of seasonal data than data after seasonal adjustment, due to the number of questions that arise in the analysis of seasonal data. For instance, there are questions as to the stationarity or otherwise of seasonal dynamics, whether capturing seasonality requires the use of periodic (seasonally-varying coefficient) models and whether there are nonlinear seasonal/business cycle interactions. According to our results, almost all forecast combination methods deliver improved forecast performance, on average, compared with individual methods. Nevertheless, the combination methods that produce the most accurate forecasts identify the best forecasting models and base the combination on these. Indeed, a simple average of the best five forecasting models for a particular horizon performs well, especially when forecasting one month ahead. Nevertheless, we find that better combinations can usually be found by weighting the forecasts using information from the root mean-square prediction error for earlier periods.

Our results relating to the use of more complex methods of handling seasonality are mixed, in the sense that nonlinear models here deliver poor forecast performance, whereas a parsiminous parameterisation of a periodic model provides the most accurate forecasts for industrial production in the Euro Area and some other countries at short horizons. This indicates that future research may examine further the extent to which the imposition of restrictions can improve the forecast performance of such methods.

References

Armstrong, H. (1989): "Combining Forecasts: The End of the Beginning or the Beginning of the End?", *International Journal of Forecasting,* 5, 585-588.

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

Beaulieu, J.J., and J.A. Miron (1991): "The Seasonal Cycle in US Manufacturing", *Economics Letters*, 37, 115-118.

Boswijk, H.P., and P.H. Franses (1996): "Unit Roots in Periodic Autoregressions", *Journal of Time Series Analysis*, 17, 221-245.

Box, G.E.P., and G.M. Jenkins (1970): *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

CECCHETTI, S.G., AND A.K. KASHYAP (1996): "International Cycles", *European Economic Review,* 40, 331-360.

CHAN, K.S. (1990): "Testing for Threshold Autoregression", *The Annals of Statistics*, 18, 1886-1894.

CLEMEN, R.T. (1989): "Combining Forecasts: a Review and Annotated Bibliography", *International Journal of Forecasting*, 5, 559-583.

CLEMENTS, M.P., AND J. SMITH (1997): "The Performance of Alternative Forecasting Methods for SETAR Models", *International Journal of Forecasting*, 13, 463-475.

CLEMENTS, M.P., P.H. FRANSES, J. SMITH AND D. VAN DIJK (2003): "On SETAR Nonlinearity and Forecasting", *Journal of Forecasting*, 22, 359-375.

DAVIES, R.B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative", *Biometrika*, 64, 247-254.

DIEBOLD, F.X. (1988): "Serial Correlation and the Combination of Forecasts", *Journal of Business and Economic Statistics* 6, 105-111.

DIEBOLD, F.X., AND J.A. LOPEZ (1996): "Forecast Evaluation and Combination", in *Handbook of Statistics*, ed. by G.S. Maddala and C.R. Rao. North-Holland, Amsterdam.

DIEBOLD, F.X., AND P. PAULY (1987): "Structural Change and the Combination of Forecasts", *Journal of Forecasting,* 6, 21-40.

DIEBOLD, F.X. (2001): *Elements of Forecasting.* Thomson Learning, Grawfordville.

FRANSES, P.H. (1996): *Periodicity and Stochastic Trends in Economic Time Series.* Oxford University Press, Oxford.

GHYSELS, E., AND D.R. OSBORN (2001): *The Econometric Analysis of Seasonal Time Series.* Cambridge University Press, Cambridge.

GHYSELS, E., D.R. OSBORN AND P.M.M. RODRIGUES (2006): „Forecasting Seasonal Time Series", in *Handbook of Economic Forecasting*, Vol. I, ed. by G. Elliott, C.W.J. Granger and A. Timmerman. Elsevier, Amsterdam.

GOMEZ, V., AND A. MARAVALL (1996): "Programs TRAMO and SEATS: Instruction for the User, Beta Version", Mimeo.

GRANGER, C.W.J., AND R. RAMANATHAN (1984): "Improved Methods of Combining Forecasting", *Journal of Forecasting*, 3, 197-204.

HANSEN, B.E. (1996): "Inference when a Nuisance Parameter is Not Identified Under the Null Hypothesis", *Econometrica*, 64, 413-430.

HANSEN, B.E. (1997): "Inference in TAR Models", *Studies in Nonlinear Dynamics and Econometrics*, 2, 1-14.

HENDRY D.F., AND M.P. CLEMENTS (2002): "Pooling of Forecasts", *Econometric Journal* 5, 1-26.

HIBON, M., AND T. EVGENIOU (2005): "To Combine or not to Combine: Selecting Among Forecasts and their Combinations", *International Journal of Forecasting*, 21, 15-24.

HYLLEBERG, S., C. JORGENSEN, AND N.K. SORENSEN (2003): "Seasonalily in Macroeconomic Time Series", *Empirical Economics,* 18, 321-335.

MATAS-MIR, A., AND D.R. OSBORN (2004): "Does Seasonality Change over the Business Cycle? An Investigation using Monthly Industrial Production Series", *European Economic Review,* 48, 1309-1332.

NEWBOLD, P., AND D. I. HARVEY (2002): "Forecast Combining and Encompassing", in *A Companion to Economic Forecasting*, ed. by M.P. Clements and D.F. Hendry. Blackwell, Oxford.

NG, S., AND P. PERRON (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag", *Journal of American Statistical Society*, 90, 268-281.

OSBORN, D.R. (1990): "A Survey of Seasonality in UK Macroeconomic Variables", *International Journal of Forecasting*, 6, 327-336.

OSBORN, D.R. (1991): "The Implications of Periodically Varying Coefficients for Seasonal Time-Series Processes", *Journal of Econometrics*, 48, 373-84.

OSBORN, D.R. (1993): "Discussion: Seasonal Cointegration", *Journal of Econometrics*, 55, 299-303.

OSBORN, D.R., AND J.P. SMITH (1989): "The Performance of Periodic Autoregressive Models in Forecasting Seasonal UK Consumption", *Journal of Business and Economic Statistics*, 7, 1117-27.

RODRIGUES, P.M.M., AND P.M.D.C. GOUVEIA (2004): "An Application of PAR Models for Tourism Forecasting", *Tourism Economics*, 10, 281-303.

STOCK, J.H., AND M.W. WATSON (1999): "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", in *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, ed. by R.F. Engle and H. White. Oxford University Press, Oxford.

STOCK, J.H. AND M.W. WATSON (2004): "Combination Forecasts of Output Growth in a Seven-Country Data Set", *Journal of Forecasting*, 23, 405-430.

TERUI, N., AND H.K. VAN DIJK (2002): "Combined Forecasts From Linear and Nonlinear Time Series Models", *International Journal of Forecasting* 18, 421-438.

TIAO, G.C., AND M.R. GRUPE (1980): "Hidden Periodic Autoregressive-Moving Average Models in Time Series Data", *Biometrika*, 67, 365-73.

TSAY, R.S. (1989): "Testing and Modeling Threshold Autoregressive Processes", *Journal of the American Statistical Association,* 84, 231-240.

Table 6: Ranking of Top Ten Forecasting Models for Each Country

	Austria			Canada			Denmark			Euro Area		
Rank	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8
1	M11	M12	M6	M1	C11	M1	C10	C9	C10	M11	M11	M11
2	C1	M11	M7	C1	C8	C14	C9	C10	C5	M7	M7	C9
3	C3	M6	M12	C3	M7	C12	C7	M4	C4	M12	C10	C14
4	C17	M7	M8	M11	M12	C17	C8	C17	C15	C16	M12	C7
5	M12	C17	C10	C17	M1	C7	C6	C7	C6	C15	M10	C6
6	M6	C16	C16	C16	C14	C16	C5	C6	C7	C4	C15	C16
7	C16	C4	C4	C14	C9	C6	C11	C16	C16	C17	C4	C5
8	C15	C15	C15	C4	C12	C9	C15	C5	C17	C5	C16	C4
9	C4	C5	C17	C15	C7	C5	C4	C4	C14	M10	C17	C15
10	C14	C14	C5	C8	C16	C15	C16	C15	M4	C10	C5	C17

	Finland			France			Germany			Greece		
Rank	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8
1	M17	M17	M17	M6	C18	C18	M17	M11	M11	M17	M10	M10
2	C10	C15	M1	C8	M6	M6	C3	M7	C9	C8	C10	C11
3	C13	C4	C4	C11	C8	C8	C8	C11	M7	C10	C8	C18
4	C11	M1	C15	C18	M12	C11	C17	C15	C17	C11	C13	C10
5	C7	C16	C5	C14	C9	M10	C16	C4	C16	M5	C15	C15
6	C9	C17	C6	C7	C11	M7	C1	C16	C7	C7	C4	C4
7	C8	C5	C14	C9	C7	C7	C4	C17	M12	C6	C5	C16
8	C2	C14	C16	C6	C6	C9	C15	C10	C6	C13	C16	C8
9	M4	C6	C7	C5	C10	C6	C5	C5	C5	C5	C6	C5
10	M10	C7	C17	C4	C5	C5	C14	C8	C14	C15	C17	C17

	Hungary			Italy			Japan			Luxembourg		
Ranking	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8
1	C10	M1	M1	C18	M1	M3	M12	C2	M1	M9	C15	M16
2	C3	C6	C9	M5	M17	C2	C5	C11	C10	C1	C4	M14
3	C17	C2	C10	C14	C8	M2	C6	C5	C17	C3	C14	M1
4	C15	C5	C15	C11	C11	C13	C15	C6	C16	C15	C16	C13
5	C4	C10	C4	C8	C2	M1	C4	C4	C15	C4	C9	M17
6	C16	C7	C5	C4	C13	C14	C7	C15	C4	C18	C5	M15
7	C5	C15	C6	C15	C12	C12	C16	C14	C5	C16	M16	C2
8	C1	C4	C7	M4	C7	C17	C10	C16	C2	C5	C6	C17
9	C6	C17	C17	C16	C10	C16	C17	C7	C6	C17	C17	M5
10	C7	C16	C16	C5	C6	C9	C8	C17	C14	C6	C10	C10

Table 7: Ranking of Top Ten Forecasting Models for Each Country (Cont.)

	Netherlands			Portugal			Spain			Sweden		
Rank	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8	h=1	h=3	h=8
1	M11	C2	M4	M5	M4	M4	C17	C10	C8	M11	M12	M6
2	C17	C13	M3	M17	M1	M1	C4	C15	C9	C1	M11	M7
3	C16	M11	M2	M4	M11	M12	C15	C4	C7	C3	M6	M12
4	C4	C12	C9	C10	M12	M11	C16	C5	M11	C17	M7	M8
5	C15	M3	C8	C15	C10	C10	C5	C8	C6	M12	C17	C10
6	C3	C9	M11	C4	M17	C13	C8	C6	C5	M6	C16	C16
7	C10	C8	C11	C16	C4	C2	C6	C16	C15	C16	C4	C4
8	C5	C14	C7	C13	C15	C14	C9	C7	C4	C15	C15	C15
9	C13	C7	C17	C17	C17	C17	C7	C17	C16	C4	C5	C17
10	C1	C6	C16	C2	C16	C16	C10	C14	C10	C14	C14	C5

	UK			USA		
Rank	h=1	h=3	h=8	h=1	h=3	h=8
1	C18	C18	M7	C11	M6	M11
2	M12	M12	M8	M6	M7	C8
3	M6	M7	M12	C8	C11	M7
4	C8	M8	M6	C9	M10	M1
5	M10	C9	C10	C2	M11	C11
6	M11	C14	M10	C14	C8	C14
7	C14	C7	C8	C13	C9	C12
8	C9	C6	C14	M1	C14	C9
9	C7	C5	C4	C3	C12	C7
10	C10	C4	C15	C12	C13	C2

Table 7: Average Rank, Average RMSPE and Scaled Average RMSPE

	h=1			h=3			h=8		
Model	Rank	RMSPE	M$_i$/M1	Rank	RMSPE	M$_i$/M1	Rank	RMSPE	M$_i$/M1
M1	23.2	0.031	1.000	18.2	0.036	1.000	15.1	0.048	1.000
M2	24.6	0.031	1.014	24.4	0.037	1.055	20.1	0.049	1.053
M3	25.2	0.031	1.017	22.4	0.037	1.047	20.3	0.050	1.076
M4	19.6	0.029	0.948	20.2	0.036	1.000	19.7	0.050	1.077
M5	23.2	0.032	1.048	30.1	0.041	1.178	28.6	0.056	1.202
M6	20.6	0.030	0.972	20.9	0.038	1.030	20.2	0.052	1.111
M7	25.1	0.032	1.010	17.8	0.038	1.028	16.8	0.049	1.069
M8	29.0	0.034	1.092	27.4	0.041	1.119	20.4	0.050	1.062
M9	33.4	0.772	28.03	35.4	0.652	20.10	35.1	0.927	20.88
M10	21.3	0.030	0.965	22.6	0.037	1.028	20.2	0.050	1.078
M11	17.9	0.030	0.964	16.2	0.036	0.991	17.4	0.049	1.043
M12	19.4	0.030	0.979	17.3	0.036	1.012	16.9	0.049	1.050
M13	33.8	0.047	1.464	35.6	0.259	8.741	35.5	0.724	18.99
M14	28.1	0.032	1.035	27.1	0.039	1.082	26.7	0.054	1.152
M15	27.5	0.032	1.026	27.7	0.039	1.087	27.5	0.055	1.186
M16	27.2	0.032	1.025	25.6	0.038	1.077	26.5	0.054	1.176
M17	17.6	0.029	0.939	18.7	0.035	0.985	21.9	0.050	1.079
	h=1			h=3			h=8		
Method	Rank	RMSPE	C$_i$/M1	Rank	RMSPE	C$_i$/M1	Rank	RMSPE	C$_i$/M1
C1	21.2	0.062	2.205	29.2	0.062	1.874	32.3	0.097	2.273
C2	15.8	0.029	0.924	12.8	0.034	0.951	14.4	0.047	0.988
C3	19.9	0.058	2.033	27.7	0.058	1.743	31.1	0.090	2.096
C4	9.3	0.028	0.906	8.3	0.034	0.933	9.3	0.046	0.968
C5	10.4	0.028	0.906	8.8	0.034	0.934	9.7	0.046	0.969
C6	10.9	0.028	0.906	9.4	0.034	0.935	10.1	0.046	0.969
C7	10.8	0.028	0.906	10.6	0.034	0.936	10.6	0.046	0.971
C8	10.4	0.028	0.909	11.8	0.034	0.940	13.8	0.047	0.992
C9	13.8	0.029	0.919	12.3	0.034	0.944	12.4	0.047	0.982
C10	9.9	0.028	0.904	9.7	0.034	0.934	9.3	0.046	0.966
C11	12.9	0.029	0.916	12.7	0.035	0.944	15.7	0.048	1.010
C12	17.1	0.029	0.929	15.1	0.035	0.957	17.2	0.048	1.002
C13	13.9	0.029	0.918	13.9	0.034	0.951	15.7	0.047	0.991
C14	12.3	0.028	0.912	11.2	0.034	0.938	11.7	0.047	0.975
C15	9.3	0.028	0.906	8.2	0.034	0.933	9.6	0.046	0.968
C16	9.3	0.028	0.904	9.4	0.034	0.934	9.1	0.046	0.968
C17	9.0	0.028	0.904	10.8	0.034	0.935	9.3	0.046	0.968
C18	21.6	0.032	1.025	23.7	0.039	1.090	21.9	0.058	1.255

Note: Scaled average RMSPE is relative to RMSPE of M1, and is denoted M_i/M_1 or C_j/M_1 ($i = 2, ..., 17$, $j = 1, ..., 17$).