The Forecasting Performance of Various Models for Seasonality and Nonlinearity for Quarterly Industrial Production[*]

Philip Hans Franses[†]
Econometric Institute
Erasmus University Rotterdam

Dick van Dijk[‡]
Econometric Institute
Erasmus University Rotterdam

Abstract

Seasonality often accounts for the major part of quarterly or monthly movements in detrended macro-economic time series. In addition, business cycle nonlinearity is a prominent feature of many such series too. A forecaster can nowadays consider a wide variety of time series models which describe seasonal variation and regime-switching behaviour. In this paper we examine the forecasting performance of various models for seasonality and nonlinearity using quarterly industrial production series for 17 OECD countries. We find that forecasting performance varies widely across series, across forecast horizons and across seasons. However, in general, linear models with fairly simple descriptions of seasonality outperform at short forecast horizons, whereas nonlinear models with more elaborate seasonal components dominate at longer horizons.

Key words: forecasting; nonlinearity; seasonality; industrial production.

---

1 Introduction

Seasonality often accounts for a major part of quarterly or monthly movements in somehow detrended macro-economic time series. While the interest in explicitly modelling these seasonal patterns has been steadily increasing in recent years, no consensus has been reached as to which model yields the most appropriate description. Among the most popular models of seasonality are (1) models with a constant (deterministic) seasonal pattern, (2) seasonal unit root models, (3) models in first and seasonal differences, and (4) periodic models. A number of studies have considered the relative forecasting performance of these models, see, for example, Osborn (2001) and Franses and Paap (2001) for reviews.

For macroeconomic time series variables such as output and unemployment, another stylized feature is business cycle asymmetry. For example, the dynamic behaviour of unemployment rates is generally characterized by steep increases during recessions and slow(er) declines during expansions, see Skalin and Teräsvirta (2001), among others. Various nonlinear time series models have been considered to describe asymmetry related to the business cycle, such as the Markov-switching model (Hamilton, 1989), the threshold autoregressive model (Tong, 1990; Potter, 1995), and the smooth transition autoregressive (STAR) model (Chan and Tong, 1986; Teräsvirta and Anderson, 1992). Franses and Van Dijk (2000) provide a review of these and other regime-switching models. In empirical applications of these models it is often reported that, even though the nonlinear model appears to describe certain characteristics of the time series at hand much better than a linear model, the forecasting performance of the linear model is no worse or even better than that of the nonlinear model, see de Gooijer and Kumar (1992), among others. This may be due to the nonlinear features not showing up during the forecast period, the use of inappropriate measures of forecast performance (Clements and Smith, 1999), or simply the fact that the supposed nonlinearity is spurious and is established by neglected heteroskedasticity, structural breaks or outliers.

In any case, no matter which model that somehow includes descriptions of seasonality and nonlinearity is put forward, it is always relevant to examine how this model performs out-of-sample. Therefore, in this paper, we consider the forecasting performance of five linear models for seasonality for quarterly industrial production series from 17 OECD countries. In addition, we consider three nonlinear models which also include regime-switching behaviour of the smooth transition type.

Our findings can be summarized as follows. We find that relative forecast performance varies widely across series, across forecast horizons and across seasons, although some interesting patterns emerge. In particular, linear models with fairly

simple descriptions of seasonality outperform at short forecast horizons, whereas nonlinear models with more elaborate seasonal components dominate at longer horizons. Additionally, a simple unweighted average of the forecasts from all individual models turns out to be a robust forecasting device, in the sense that it performs well at both short and long horizons.

The outline of our paper is as follows. In Section 2, we review the models for seasonality and nonlinearity under scrutiny. For each model we also outline the specification procedure used in the empirical application. In Section 3, we discuss the results from our forecast comparison for the quarterly industrial production series. Section 4 concludes with some remarks.

2 Models of seasonality and nonlinearity

In this section we review various time series models of seasonality and nonlinearity which we aim to compare in our forecasting exercise. We describe the different models having in mind the application to a quarterly time series. Generalizations to variables observed at other frequencies are straightforward. For all models, except for the periodic autoregressive model in Section 2.4, we a priori assume a unit root at the zero frequency and, hence, we detrend the series by first differencing. As we also transform the data by taking logarithms, we subsequently analyze quarterly growth rates.

2.1 A linear model with stable deterministic seasonality

The first (and presumably simplest) model of seasonality posits that the seasonal pattern in the time series y_t is constant over time and that it can be captured by seasonal dummies $D_{s,t}$, $s = 1, \ldots, 4$, defined as $D_{s,t} = 1$ if time t corresponds to season s and $D_{s,t} = 0$ otherwise. This model, labelled as AR in the following, reads as

$$\phi(L)\Delta_1 y_t = \delta_1 D_{1,t} + \delta_2 D_{2,t} + \delta_3 D_{3,t} + \delta_4 D_{4,t} + \varepsilon_t, \qquad (1)$$

where $\phi(L)$ is a polynomial of order p in the lag operator L with all roots outside the unit circle, Δ_1 is the differencing operator defined as $\Delta_k y_t = y_t - y_{t-k}$ for all integer $k \neq 0$, and where $\varepsilon_t \sim \mathsf{NID}(0, \sigma^2)$, although we can allow for, for example, season-specific variances in the disturbances.

To specify model (1), the only decision to be made in practice concerns selecting the lag order p. In the application below we select p as the order that minimizes the Schwarz Information Criterion (BIC), subject to passing an Lagrange Multiplier (LM) test for first-to-fifth order residual autocorrelation at the 5% significance level. The maximum order is set equal to $p_{\max} = 8$.

2.2 A linear model with seasonal unit roots

The AR model (1) implies a constant seasonal pattern, whereas many macroeconomic time series seem to be characterized by seasonal patterns which evolve over time, see Hylleberg, Jørgensen and Sørensen (1993), and Canova and Hansen (1995), among many others. A popular approach to allow for time-varying seasonality is the autoregressive model for seasonally differenced data, that is,

$$\phi(L)\Delta_4 y_t = \mu + \varepsilon_t, \tag{2}$$

where $\phi(L)$ and ε_t are defined before (but p and σ^2 may of course take values different from the ones in (1)). In this model, the time series is driven by a nonstationary process in each season, while the seasons share a common drift μ. This implies that in each season the process is trending with the same rate. For example, when $\phi(L) = 1$, y_t depends on the sum of ε_t, ε_{t-4}, ... (plus an annual linear trend with slope μ), while y_{t-i}, $i = 1, 2, 3$ depends on the sum of ε_{t-i}, ε_{t-4-i}. Hence, the processes in each quarter are the sum of independent increments. This gives rise to the notion that this model allows "winter to become summer."

As $\Delta_4 = (1-L)(1+L)(1-iL)(1+iL)$ with $i^2 = -1$, the model in (2) implies unit roots at the seasonal frequencies $\pi/2$ and π, in addition to the zero frequency unit root. Hylleberg *et al.* (1990) develop statistical tests for testing for such seasonal unit roots. For simplicity, here we impose all seasonal unit roots *a priori*. The only practical decision then to be made concerns selecting the autoregressive order p. As for the model with deterministic seasonality, we select p as the order that minimizes the BIC, subject to passing an LM test for first-to-fifth order residual autocorrelation at the 5% significance level. The maximum order is set equal to $p_{\max} = 8$. We label this model as SUR below.

2.3 A SARIMA model

The SUR model (2) can be regarded as a member of the class of seasonal autoregressive integrated moving average (SARIMA) models developed by Box and Jenkins (1970). For quarterly data, the general SARIMA model can be written as

$$\phi(L)\Delta_1 \Delta_4 y_t = \mu + (1 - \theta_1 L)(1 - \theta_4 L^4)\varepsilon_t \tag{3}$$

where $|\theta_1| < 1$, $|\theta_4| < 1$. As before, $\phi(L)$ has all of its roots outside the unit circle and often it is restricted to the form $\phi(L) = (1 - \phi_1 L)(1 - \phi_4 L^4)$. The "airline model" of Box and Jenkins (1970) is a special case when $\phi(L) = 1$ and $\mu = 0$. Here we leave $\phi(L)$ (and μ) unrestricted, and determine the autoregressive order in the same manner as for the linear model with deterministic seasonality. Notice that

the SARIMA model assumes seasonal unit roots at frequencies $\pi/2$ and π and two unit roots at the zero frequency. Empirical estimates of θ_1 and θ_4 are often close to unity however, giving rise to more regular seasonal patterns than consideration of the autoregressive part of the model alone would suggest.

2.4 A periodic autoregressive model

An entirely different approach to allow for flexible seasonal patterns is offered by periodic autoregressive (PAR) models, which suppose that not only the intercept (and trend if present) is seasonal, but also that the autoregressive parameters vary with the season. An unrestricted PAR model with seasonal dummies and seasonal trends is given by

$$y_t = \mu_s + \tau_s T_t + \phi_{1,s} y_{t-1} + \ldots + \phi_{p,s} y_{t-p} + \varepsilon_t, \tag{4}$$

where $T_t = [(t-1)/4] + 1$. Models of this sort have become prominent in economics over the last decade following the seminal studies by Osborn (1988) and Osborn and Smith (1989), see Franses (1996) and Franses and Paap (2002) for reviews.

Franses and Paap (2001) suggest that the evidence of improved forecasting performance from using PAR models is rather mixed. They review the role of unit roots and deterministic terms (trends as well as intercepts) in PAR models, and suggest that an appropriate treatment of these aspects of the model can result in more accurate forecasts. Here we closely follow the specification strategy outlined in Paap and Franses (1999). In the first step, we determine the appropriate lag order in the unrestricted model (4), using the BIC in conjunction with an LM test for first-to-fifth order autocorrelation. In the second step, we test for periodic integration, which boils down to testing the restriction $\alpha_1\alpha_2\alpha_3\alpha_4 = 1$ in the equivalent representation

$$(y_t - \alpha_s y_{t-1} - \mu_s^* - \tau_s^* T_t) = \sum_{i=1}^{p-1} \beta_{i,s}(y_{t-i} - \alpha_s y_{t-1-i} - \mu_{s-i}^* - \tau_{s-i}^* T_{t-i}) + \varepsilon_t, \tag{5}$$

where μ_s^*, τ_s^* and $\beta_{i,s}$ are nonlinear functions of the μ_s, τ_s and $\phi_{i,s}$ parameters in (4), and where $\mu_{s-4k}^* = \mu_s^*$ and $\tau_{s-4k}^* = \tau_s^*$ for $k = 1, 2, \ldots$.

In general, periodic models with unrestricted seasonal linear trends as in (4) can generate data with diverging seasonal trends. Similarly, the inclusion of linear deterministic trends in the periodically integrated autoregression (PIAR) (5) with $\alpha_1\alpha_2\alpha_3\alpha_4 = 1$ implies the presence of seasonal quadratic trends in y_t. Both may not be plausible for economic time series. In the third step, we therefore test several restrictions on the seasonal intercepts and trend parameters which restrict the implied trending behaviour of the time series in different ways. On the one hand, if periodic integration is not rejected, we use a sequence of likelihood ratio tests to

4

test restrictions on the intercepts and trend parameters in (5) which imply (i) the absence of quadratic trends, (ii) the presence of a common (seasonal) linear trend, and (iii) the absence of linear trends, see Paap and Franses (1999) for details. On the other hand, if periodic integration is rejected, we proceed with the representation

$$(y_t - \mu_s^{**} - \tau_s^{**}T_t) = \sum_{i=1}^{p} \phi_{i,s}(y_{t-i} - \mu_{s-i}^{**} - \tau_{s-i}^{**}T_{t-i}) + \varepsilon_t, \qquad (6)$$

where μ_s^{**}, τ_s^{**} are again nonlinear functions of μ_s, τ_s and $\phi_{i,s}$ and $\mu_{s-4k}^{**} = \mu_s^{**}$ and $\tau_{s-4k}^{**} = \tau_s^{**}$ for $k = 1, 2, \ldots$. In (6), we test for (i) a common linear seasonal deterministic trend ($\tau_1^{**} = \tau_2^{**} = \tau_3^{**} = \tau_4^{**}$), and (ii) for the absence of linear deterministic trends ($\tau_1^{**} = \tau_2^{**} = \tau_3^{**} = \tau_4^{**} = 0$).

2.5 A linear model with smoothly changing deterministic seasonality

It is well known that a change in the deterministic trend properties of a time series y_t is easily mistaken for the presence of a unit root. In a similar vein, if a change in the deterministic seasonal pattern in models such as (1) goes undetected, one might well end up imposing seasonal unit roots as in (2). Ghysels (1994), Smith and Otero (1997), Franses, Hoek and Paap (1997) and Franses and Vogelsang (1998) document that neglected deterministic seasonal mean shifts in (1) bias the outcomes of the tests of Hylleberg *et al.* (1990) towards non-rejection of the seasonal unit root hypothesis. Intuitively, it may be difficult to distinguish a seasonal unit root model from a model with changing deterministic seasonality. Such a distinction may be important however from a forecasting perspective, see Clements and Hendry (1997) and Paap, Franses and Hoek (1997) for discussion.

Changes in deterministic seasonal patterns usually are modelled by means of one-time abrupt and discrete changes, see the aforementioned studies for examples. However, seasonal patterns often change slowly over time due to changes in technology, institutions and tastes, for example. This suggests that a plausible description of time-varying seasonal patterns is an extension of (1) with smoothly changing deterministic seasonals, that is,

$$\phi(L)\Delta_1 y_t = \sum_{s=1}^{4} \delta_{1,s} D_{s,t}(1 - G(t; \gamma, c)) + \sum_{s=1}^{4} \delta_{2,s} D_{s,t} G(t; \gamma, c) + \varepsilon_t, \qquad (7)$$

where $G(t; \gamma, c)$ is the logistic function

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)\}}, \qquad \gamma > 0. \qquad (8)$$

As s_t increases, the logistic function changes monotonically from 0 to 1, with the change being symmetric around the location parameter c, as $G(c-z;\gamma,c) = 1-G(c+z;\gamma,c)$ for all z. The slope parameter γ determines the smoothness of the change. As $\gamma \to \infty$, the logistic function $G(s_t;\gamma,c)$ approaches the indicator function $\mathtt{I}[s_t > c]$, whereas if $\gamma \to 0$, $G(s_t;\gamma,c) \to 0.5$ for all values of s_t. Hence, by taking $s_t = t$ in (8), both the model with instantaneous change and the AR model (1) with no change in the seasonal pattern are nested in (7). In the terminology of Franses and Vogelsang (1998), the model in (7) takes an "intermediate" position in between deterministic seasonality as in (1) and stochastic trend seasonality as in (2).

In specifying (7), again the only choice to be made is selecting the appropriate order of the autoregressive lag polynomial $\phi(L)$. As for the other models, we will select p by minimizing the BIC subject to non-rejection of first-to-fifth order residual autocorrelation by an appropriate LM test at the 5% significance level. We will call this model the TV-AR model.

2.6   A nonlinear model with deterministic seasonality

As discussed in the introduction, apart from prominent seasonal patterns, macroeconomic time series such as output and unemployment also exhibit asymmetry related to the business cycle. To capture this feature, we extend three of the models discussed above to allow for such regime-switching behaviour in the dynamics.

First, we consider a smooth transition autoregressive (STAR) model, put forward in Teräsvirta (1994), augmented with a deterministic seasonal pattern for the first-differenced time series, that is,

$$\Delta_1 y_t = \delta_1 D_{1,t}^* + \delta_2 D_{2,t}^* + \delta_3 D_{3,t}^* + (\mu_1 + \phi_{1,1}\Delta_1 y_{t-1} + \cdots + \phi_{1,p}\Delta_1 y_{t-p})(1-G(\Delta_4 y_{t-d};\gamma,c))$$
$$+ (\mu_2 + \phi_{2,1}\Delta_1 y_{t-1} + \cdots + \phi_{2,p}\Delta_1 y_{t-p})G(\Delta_4 y_{t-d};\gamma,c) + \varepsilon_t, \quad (9)$$

where $D_{s,t}^* = D_{s,t} - D_{4,t}$, $s = 1,2,3$, and $G(\Delta_4 y_{t-d};\gamma,c)$ with $d > 0$ is a logistic function as in (8).

The use of $\Delta_4 y_{t-d}$ as the variable determining the regime-switches in (9) is motivated by the fact that this variable tracks the business cycle quite closely for our quarterly industrial production series. Because the logistic function $G(s_t;\gamma,c)$ is a monotonic transformation of s_t, the regimes associated with $G(\Delta_4 y_{t-d};\gamma,c) = 0$ and $G(\Delta_4 y_{t-d};\gamma,c) = 1$ will roughly correspond with recessions and expansions, respectively, see Skalin and Teräsvirta (2001) for additional discussion. Notice that (9) can be considered as a generalization of the AR model with stable seasonality (1), allowing for different autoregressive dynamics in expansions and recessions.

By specifying the deterministic terms in (9) in terms of an intercept μ_i, $i = 1, 2$ and "centered" seasonal dummies $D_{s,t}^*$, we allow the overall growth rate of y_t to vary

according to the phase of the business cycle, but restrict the seasonal pattern to be the same in expansions and recessions. Even though Canova and Ghysels (1994) and Franses (1996, pp. 86-87) conclude that the business cycle influences the seasonal cycle of US output, van Dijk, Strikholm and Teräsvirta (2001) find no such evidence for the quarterly industrial production series of the G7 countries, which are among the series we analyze in this paper.

To specify the model in (9), we follow the modelling cycle for STAR models designed in Teräsvirta (1994, 1998). We start from the AR model (1) and test linearity against the alternative given in (9) using the LM-type tests developed in Luukkonen, Saikkonen and Teräsvirta (1988). We test linearity against alternatives with $\Delta_4 y_{t-d}$ for $d = 1, \ldots, 4$, and select the appropriate value of d as the one which minimizes the p-value of the corresponding test statistic. After fixing the transition variable, we re-specify the autoregressive order p in (9) by minimizing the BIC subject to passing an LM-type test of no first-to-fifth order residual autocorrelation at the 5% significance level. We will call the model in (9) the STAR model.

2.7 A nonlinear model with smoothly changing deterministic seasonality

In identical fashion, we extend the linear model with smoothly changing deterministic seasonality (7) to allow for business cycle asymmetry. This results in the model

$$
\begin{aligned}
\Delta_1 y_t = & [\mu_1(1 - G(\Delta_4 y_{t-d}; \gamma_2, c_2)) + \mu_2 G(\Delta_4 y_{t-d}; \gamma_2, c_2)](1 - G(t; \gamma_1, c_1)) \\
& + [\mu_3(1 - G(\Delta_4 y_{t-d}; \gamma_2, c_2)) + \mu_4 G(\Delta_4 y_{t-d}; \gamma_2, c_2)]G(t; \gamma_1, c_1) \\
& + \sum_{s=1}^{3} \delta_{1,s} D_{s,t}^* (1 - G(t; \gamma_1, c_1)) + \sum_{s=1}^{3} \delta_{2,s} D_{s,t}^* G(t; \gamma_1, c_1) \\
& + (\phi_{1,1}\Delta_1 y_{t-1} + \cdots + \phi_{1,p}\Delta_1 y_{t-p})(1 - G(\Delta_4 y_{t-d}; \gamma_2, c_2)) \\
& + (\phi_{2,1}\Delta_1 y_{t-1} + \cdots + \phi_{2,p}\Delta_1 y_{t-p})G(\Delta_4 y_{t-d}; \gamma_2, c_2) + \varepsilon_t, \quad (10)
\end{aligned}
$$

where again $D_{s,t}^* = D_{s,t} - D_{4,t}$, $s = 1, 2, 3$. As in the specification in (9), the seasonal pattern is allowed to vary over time only and is assumed to be identical in expansions and recessions at all times. The model in (10) can be considered as a special case of the time-varying STAR (TV-STAR) model discussed in detail in Lundbergh, Teräsvirta and van Dijk (2000).

Specification of the TV-STAR model proceeds as follows. Starting from the TV-AR model (7), linearity is tested against the alternative (10) with $\Delta_4 y_{t-d}$ for $d = 1, \ldots, 4$. The value of d is selected as the one which minimizes the p-value of the corresponding test statistic. After fixing the transition variable, we re-specify

the autoregressive order p in (10) by minimizing the BIC subject to an LM test of no first-to-fifth order residual autocorrelation at the 5% significance level.

2.8 A nonlinear model with seasonal unit roots

Our final model combines the seasonal unit root model (2) with smooth transition type regime-switching dynamics, that is,

$$\Delta_4 y_t = (\mu_1 + \phi_{1,1}\Delta_4 y_{t-1} + \cdots + \phi_{1,p}\Delta_4 y_{t-p})(1 - G(\Delta_4 y_{t-d}; \gamma, c))$$
$$+ (\mu_2 + \phi_{2,1}\Delta_4 y_{t-1} + \cdots + \phi_{2,p}\Delta_4 y_{t-p})G(\Delta_4 y_{t-d}; \gamma, c) + \varepsilon_t. \quad (11)$$

This model has been applied successfully to a subset of the quarterly industrial production series considered here on a shorter sample by Teräsvirta and Anderson (1992).

The model is specified by testing the SUR model (2) against the alternative given in (11) to determine the appropriate value of the delay parameter d, using the same decision rule as for the previous two models. Similarly, after selecting d, the autoregressive order in (11) is re-specified by minimizing BIC subject to an LM-type test of no residual autocorrelation at first-to-fifth order. We will label this model as the SUR-STAR model. In the next section, we will evaluate this model and the other seven models discussed above on their forecasting merits.

3 Empirical results

Our data set consists of quarterly seasonally unadjusted industrial production volume indexes for 17 OECD countries, including Austria, Belgium, Canada, Finland, France, Germany, Greece, Italy, Japan, Luxembourg, The Netherlands, Norway, Portugal, Spain, Switzerland, the United Kingdom (UK) and the United States (US). The data are taken from the OECD *Main Economic Indicators*. The sample period runs from 1960.1 until 1998.4. For Canada and Spain the series is available only from 1961.1, while for Greece the series starts in 1962.1. Obvious outliers in 1963.1 and 1968.2 for France and in 1969.4 for Italy are replaced by the average of the index values in the same quarter of the previous and the following year. The data are transformed by taking logarithms.

We consider the forecasting performance of the eight models discussed in the previous section over the period 1989.1-1998.4 based on specifying and estimating the models on an expanding window of data, starting with 1960.1-1988.4. For each window, all models are specified and estimated afresh as outlined above and 1- to 12-steps ahead forecasts for the log-level of the series are generated, rendering a

total of 40 1-step ahead forecasts, 39 2-steps ahead forecasts, up to 29 12-steps ahead forecasts. To obtain the forecasts of the models which include a nonlinear component, we use a Monte Carlo forecasting method with resampling from the residuals, see Lundbergh and Teräsvirta (2001) for detailed discussion of forecasting with smooth transition models.

Table 1 contains ratios of the root mean squared prediction error [RMSPE] for the different models relative to the (arithmetic) average RMSPE across all eight models for horizons $h = 1$, 4, 8 and 12, which in general will be the horizons of most interest. Ranks of the different models are given in parentheses, where the model with the smallest RMSPE is given rank 1, and so on.

At first sight it appears difficult to find any consistent patterns in the results, in the sense that the relative forecasting performance varies widely across series and across horizons. Upon closer inspection of Table 1 however, some interesting patterns emerge. In particular, linear models with a simple description of seasonality (AR, SUR and SARIMA) perform relatively well at short horizons ($h = 1$, 4), whereas (non)linear models with more elaborate seasonal components (PAR and TV-STAR in particular) perform better at longer horizons ($h = 8$, 12). This is borne out quite clearly in Table 2, which contains average ranks across countries, and the geometric mean of the RMSPE ratios shown in Table 1. For 1-step ahead forecasts, the SARIMA model is ranked highest on average and has the smallest mean relative RMSPE. Its forecasting performance deteriorates quickly though as the forecast horizon becomes longer. The SUR and AR models are the second- and third-best models at $h = 1$. Even though their (relative) performance worsens at longer horizons, these models appear to be reasonably robust forecasting devices. The most notable improvement in forecasting performance as h increases is achieved by the PAR model and by the TV-AR and TV-STAR models which allow for smoothly changing deterministic seasonality. Finally, the STAR and SUR-STAR model perform worse than average at all horizons. Hence, only allowing for nonlinearity in the autoregressive dynamics is not sufficient to capture the features of the industrial production series, at least not from a forecasting perspective.

- insert Tables 1 and 2 about here -

We further evaluate the relative forecast performance of the different models by pairwise comparison using the test statistic developed by Diebold and Mariano (1995). Let $\widehat{y}_{i,t|t-h}$ and $\widehat{y}_{j,t|t-h}$ denote two competing h-step ahead forecasts of y_t, obtained from models i and j, respectively. Diebold and Mariano (1995) propose a test of the null of equal forecast accuracy for arbitrary loss function $g(e_{i,t|t-h})$, where $e_{i,t|t-h}$ is the corresponding h-step ahead forecast error for model i, $e_{i,t|t-h} =$

$y_t - \widehat{y}_{i,t|t-h}$. The loss differential is defined as $d_t \equiv g(e_{i,t|t-h}) - g(e_{j,t|t-h})$, so that equal forecast accuracy entails $\mathsf{E}[d_t] = 0$. Given a covariance-stationary sample realization $\{d_t\}_{t=T+h}^{T+n}$ (where in our case $n = 40$ for one-step ahead forecasts), the Diebold-Mariano statistic for testing the null hypothesis of equal forecast accuracy is given by

$$\frac{\overline{d}}{\sqrt{\widehat{V(\overline{d})}}}, \tag{12}$$

where $\overline{d}$ is the sample mean loss differential $\overline{d} = (n - h + 1)^{-1} \sum_{t=T+h}^{T+n} d_t$, and where $\widehat{V(\overline{d})}$ is a consistent estimate of the asymptotic variance of $\overline{d}$ which is computed as an unweighted sum of the sample autocovariances, that is,

$$\widehat{V(\overline{d})} = \frac{1}{n - h + 1} \left(\widehat{\gamma}_0 + 2 \sum_{k=0}^{h-1} \widehat{\gamma}_k \right),$$

where

$$\widehat{\gamma}_k = \frac{1}{n - h + 1} \sum_{t=T+h+k}^{T+n} \left(d_t - \overline{d} \right) \left(d_{t-k} - \overline{d} \right),$$

assuming that h-step ahead forecasts exhibit dependence up to order $h - 1$.

Table 3 lists the number of countries for which the Diebold-Mariano test statistic rejects the null hypothesis that model i's and model j's forecast performance at horizon h as measured by MSPE (that is, (12) is computed for the quadratic loss function $g(e_{i,t|t-h}) = e_{i,t|t-h}^2$) are equal in favor of the one-sided alternative that model i's performance is superior at the 5% significance level. For example, for $h = 1$ the forecast performance of the AR model is found to be significantly better than the forecast performance of the PAR model for 4 of the 17 countries. The test results confirm the observations made above. At short horizons, the AR, SUR and SARIMA models render better forecasts than the more elaborate models for quite a few countries, whereas the reverse almost never occurs. At longer horizons ($h = 8$, 12) though, the more prolific PAR, TV-AR and TV-STAR dominate these simpler models, as well as the STAR and SUR-STAR models.

- insert Table 3 about here -

A final comparison is made in Table 4, which contains the number of countries for which the Diebold-Mariano type forecast encompassing test of Harvey, Leybourne and Newbold (1998) does not reject the null hypothesis that model i's forecast at horizon h encompasses model j's forecast at the 5% significance level. This statistic is obtained by computing (12) with $d_t = e_{i,t|t-h}(e_{i,t|t-h} - e_{j,t|t-h})$. The test results again confirm the conclusion that simpler models (AR and SARIMA) perform relatively

10

well at short horizons ($h = 1, 4$), whereas more elaborate models (PAR, TV-AR and TV-STAR) perform better at longer horizons ($h = 8, 12$). As an example, the number of series for which the forecasts from the SARIMA model encompass the forecasts from the TV-STAR declines from 14 via 10 and 4 to 3, as the forecast horizon h increases from 1 via 4 and 8 to 12 quarters ahead.

- insert Table 4 about here -

Two final observations are of interest. First, a detailed examination of the results shows that the relative forecast performance of the models varies considerably across quarters. For illustrative purposes, Table 5 contains ratios of the RMSPE for 1-step ahead forecasts per quarter for the G7 countries, relative to the RMSPE of an unweighted average of the forecast from all individual models. Second, most values of the RMSPE ratios in this table are larger than one. This demonstrates the potential gains from combining the forecasts from the different models. The varying performance of the models across seasons also suggests that when considering forecast combinations, it may be worthwhile to consider season-specific weights for the individual models. This is beyond the scope of the present paper and is left for future research. Similar patterns emerge for other forecast horizons and for other countries.

- insert Table 5 about here -

4   Concluding remarks

The main conclusion from this paper is that simpler models for seasonality yield better forecasts for short horizons, while more elaborate models perform better for longer horizons. Hence, if the aim is to forecast two to three years ahead for quarterly data, it pays off to spend some effort in constructing time series models which sufficiently capture the salient features of seasonality and nonlinearity.

A natural, though not necessarily trivial, extension of our results involves considering multivariate time series. For linear and trending data some results are available, see Löf and Franses (2001), but an extension to multivariate models for nonlinear data still needs to be worked out. This seems to be a relevant topic for further research, as the results presented in this paper reassuringly show that more elaborate models, which allow for a better understanding of the data in terms of changes in seasonal patterns and of business cycle asymmetry, also yield better long-run forecasts.

References

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

Canova, F. and E. Ghysels (1994), Changes in seasonal patterns: are they cyclical?, *Journal of Economic Dynamics and Control* **18**, 1143–1171.

Canova, F. and B.E. Hansen (1995), Are seasonal patterns constant over time? A test for seasonal stability, *Journal of Business & Economic Statistics* **13**, 237–252.

Chan, K.S. and H. Tong (1986), On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* **7**, 179–190.

Clements, M.P. and D.F. Hendry (1997), An empirical study of seasonal unit roots in forecasting, *International Journal of Forecasting* **13**, 341–356.

Clements, M.P. and J. Smith (1999), A Monte Carlo study of the forecasting performance of empirical SETAR models, *Journal of Applied Econometrics* **14**, 123–142.

de Gooijer, J.G. and K. Kumar (1992), Some recent developments in non-linear time series modelling, testing and forecasting, *International Journal of Forecasting* **8**, 135–156.

Diebold, F.X. and R.S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253–263.

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

Franses, P.H., H. Hoek and R. Paap (1997), Bayesian analysis of seasonal unit roots and seasonal mean shifts, *Journal of Econometrics* **78**, 359–380.

Franses, P.H. and R. Paap (2001), Forecasting with periodic autoregressive time series model, in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell Publishers, to appear.

Franses, P.H. and R. Paap (2002), *Periodic Time Series Models*, Oxford: Oxford University Press, to appear.

Franses, P.H. and D. van Dijk (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

Franses P.H. and T.J. Vogelsang (1998), On seasonal cycles, unit roots, and mean shifts, *Review of Economics and Statistics* **78**, 231–240.

Ghysels, E. (1994), On the economics and econometrics of seasonality in economic time series, in C. Sims (ed.), *Advances in Econometrics, 6th World Congress of the Econometric Society*, Cambridge: Cambridge University Press, pp. 257–316.

Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series subject to changes in regime, *Econometrica* **57**, 357–384.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1998, Tests for forecast encompassing, *Journal of Business and Economic Statistics* **16**, 254–259.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo (1990), Seasonal integration and cointegration, *Journal of Econometrics* **44**, 215–238.

Hylleberg, S., C. Jørgensen and N.K. Sørensen (1993), Seasonality in macroeconomic time series, *Empirical Economics* **18**, 321–325.

Löf, M. and P.H. Franses (2001), On forecasting cointegrated seasonal time series, *International Journal of Forecasting*, to appear.

Lundbergh, S. and T. Teräsvirta (2001), Forecasting with smooth transition autoregressive models, in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell Publishers, to appear.

Lundbergh, S., T. Teräsvirta and D. van Dijk (2000), Time-varying smooth transition autoregressive models, Working Paper Series in Economics and Finance, No. 376, Stockholm School of Economics.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988), Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Osborn, D.R. (1988), Seasonality and habit persistence in a life cycle model of consumption, *Journal of Applied Econometrics* **3**, 255–266.

Osborn, D.R. (2001), Unit root versus deterministic representations of seasonality for forecasting, in M.P. Clements and D.F. Hendry (eds.), *A Companion to Economic Forecasting*, Oxford: Blackwell Publishers, to appear.

Osborn, D.R. and J.P. Smith (1989), The performance of periodic autoregressive models in forecasting seasonal UK consumption, *Journal of Business & Economic Statistics* **7**, 117–127.

Paap, R., P.H. Franses and H. Hoek (1997), Mean shifts, unit roots and forecasting seasonal time series, *International Journal of Forecasting* **13**, 357–368.

Paap, R. and P.H. Franses (1999), On trends and constants in periodic autoregressions, *Econometric Reviews* **18**, 271–286.

Potter, S.M. (1995), A nonlinear approach to US GNP, *Journal of Applied Econometrics* **10**, 109–125.

Skalin, J. and T. Teräsvirta (2001), Modelling asymmetries and moving equilibria in unemployment rates, *Macroeconomic Dynamics*, to appear.

Smith, J. and J. Otero (1997), Structural breaks and seasonal integration, *Economics Letters* **56**, 13–19.

Teräsvirta, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T. (1998), Modeling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Teräsvirta, T. and H.M. Anderson (1992), Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**, S119–S136.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford: Oxford University Press.

van Dijk, D., B. Strikholm and T. Teräsvirta (2001), The effects of institutional and technological change and business cycle fluctuations on seasonal patterns in quarterly industrial production series, Working Paper Series in Economics and Finance No. 429, Stockholm School of Economics.

Table 1: Out-of-sample Forecasting Performance

				Model				
Horizon	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
Austria								
1	1.02 (6)	0.92 (1)	0.99 (3)	1.02 (5)	1.04 (8)	1.01 (4)	1.03 (7)	0.98 (2)
4	0.94 (2)	0.97 (3)	1.06 (7)	1.05 (6)	1.03 (5)	0.88 (1)	1.10 (8)	0.97 (4)
8	0.93 (1)	0.95 (4)	1.12 (8)	1.12 (7)	1.02 (6)	0.94 (3)	0.99 (5)	0.93 (2)
12	0.94 (3)	0.95 (4)	1.10 (6)	1.14 (8)	0.98 (5)	0.90 (2)	0.86 (1)	1.13 (7)
Belgium								
1	1.08 (8)	0.97 (2)	0.98 (3)	0.98 (4)	0.94 (1)	1.02 (6)	1.00 (5)	1.02 (7)
4	1.02 (5)	1.03 (7)	1.03 (6)	0.98 (3)	0.94 (1)	0.96 (2)	0.99 (4)	1.05 (8)
8	1.02 (7)	1.00 (4)	1.07 (8)	1.01 (5)	0.97 (3)	1.01 (6)	0.95 (1)	0.97 (2)
12	1.02 (6)	1.00 (4)	1.15 (8)	1.00 (5)	0.94 (2)	1.04 (7)	0.89 (1)	0.95 (3)
Canada								
1	0.81 (1)	0.84 (2)	0.84 (3)	1.22 (8)	1.19 (7)	0.96 (4)	1.15 (6)	1.00 (5)
4	0.93 (2)	0.93 (3)	1.00 (5)	0.89 (1)	0.93 (4)	1.16 (8)	1.05 (6)	1.11 (7)
8	0.97 (4)	0.97 (5)	1.12 (7)	0.87 (1)	0.95 (2)	1.14 (8)	0.96 (3)	1.02 (6)
12	0.93 (2)	0.96 (5)	1.23 (8)	0.88 (1)	0.96 (4)	1.09 (7)	0.99 (6)	0.96 (3)
Finland								
1	0.99 (4)	0.99 (5)	0.89 (1)	1.06 (7)	0.96 (2)	1.11 (8)	0.97 (3)	1.03 (6)
4	0.96 (4)	0.96 (2)	1.05 (7)	0.96 (3)	0.95 (1)	1.05 (6)	1.00 (5)	1.08 (8)
8	0.96 (4)	0.94 (3)	1.15 (8)	0.91 (1)	0.93 (2)	1.05 (6)	0.98 (5)	1.08 (7)
12	0.92 (3)	0.93 (4)	1.29 (8)	0.89 (2)	0.88 (1)	1.08 (7)	0.96 (5)	1.05 (6)
France								
1	1.10 (7)	0.84 (1)	0.85 (2)	0.97 (4)	1.04 (5)	1.13 (8)	1.09 (6)	0.97 (3)
4	0.95 (2)	0.98 (4)	1.01 (5)	0.90 (1)	1.01 (6)	1.01 (7)	1.15 (8)	0.98 (3)
8	0.95 (2)	0.95 (3)	1.03 (6)	0.95 (4)	0.95 (5)	1.04 (7)	1.20 (8)	0.94 (1)
12	0.96 (5)	0.96 (4)	0.96 (6)	0.92 (2)	0.90 (1)	1.08 (7)	1.28 (8)	0.94 (3)
Germany								
1	0.95 (3)	0.91 (2)	0.90 (1)	1.18 (8)	0.98 (4)	0.99 (6)	0.98 (5)	1.11 (7)
4	0.97 (4)	0.97 (2)	1.13 (8)	1.00 (6)	0.97 (3)	0.99 (5)	1.00 (7)	0.96 (1)
8	0.98 (5)	0.99 (6)	1.24 (8)	0.97 (4)	0.91 (1)	0.96 (3)	0.95 (2)	1.01 (7)
12	0.99 (6)	0.99 (5)	1.35 (8)	0.97 (3)	0.83 (1)	0.98 (4)	0.87 (2)	1.04 (7)
Greece								
1	0.91 (1)	1.01 (7)	0.92 (2)	1.28 (8)	1.01 (6)	0.92 (3)	0.94 (4)	1.00 (5)
4	1.07 (7)	1.03 (5)	1.04 (6)	0.87 (1)	1.02 (4)	0.97 (3)	0.88 (2)	1.12 (8)
8	1.13 (7)	1.02 (5)	1.00 (4)	0.89 (2)	0.93 (3)	1.02 (6)	0.82 (1)	1.19 (8)
12	1.15 (7)	1.03 (4)	1.07 (6)	0.77 (1)	0.89 (3)	1.06 (5)	0.81 (2)	1.22 (8)
Italy								
1	0.93 (3)	1.04 (6)	0.91 (1)	1.18 (8)	0.99 (5)	0.94 (4)	0.93 (2)	1.08 (7)
4	1.00 (3)	1.02 (5)	1.07 (8)	0.86 (1)	0.94 (2)	1.07 (7)	1.02 (6)	1.01 (4)
8	1.04 (4)	1.05 (5)	1.24 (8)	0.82 (1)	0.85 (2)	1.08 (7)	0.87 (3)	1.05 (6)
12	1.09 (5)	1.11 (6)	1.28 (8)	0.80 (3)	0.74 (1)	1.08 (4)	0.77 (2)	1.11 (7)
Japan								
1	1.01 (6)	0.91 (3)	0.87 (2)	1.01 (5)	1.06 (7)	1.28 (8)	0.99 (4)	0.87 (1)
4	1.10 (8)	1.00 (4)	0.92 (2)	0.79 (1)	0.94 (3)	1.08 (6)	1.06 (5)	1.10 (7)
8	1.20 (8)	1.03 (4)	0.90 (3)	0.69 (1)	0.86 (2)	1.16 (7)	1.03 (5)	1.13 (6)
12	1.25 (7)	1.07 (5)	0.91 (4)	0.64 (1)	0.78 (2)	1.32 (8)	0.90 (3)	1.14 (6)

				Model				
Horizon	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
Luxembourg								
1	0.95 (2)	1.00 (5)	0.99 (3)	1.10 (8)	0.93 (1)	1.04 (7)	1.00 (6)	1.00 (4)
4	0.95 (2)	0.96 (4)	1.18 (8)	1.04 (7)	0.99 (6)	0.95 (1)	0.96 (3)	0.99 (5)
8	0.87 (1)	0.90 (2)	1.38 (8)	1.07 (7)	0.91 (4)	1.01 (6)	0.90 (3)	0.97 (5)
12	0.84 (4)	0.75 (1)	1.55 (8)	1.17 (6)	0.80 (3)	1.21 (7)	0.80 (2)	0.89 (5)
Netherlands								
1	0.84 (1)	0.84 (3)	0.89 (4)	1.41 (8)	1.08 (7)	0.84 (2)	1.08 (6)	1.02 (5)
4	0.90 (3)	0.99 (4)	1.02 (5)	1.17 (8)	1.10 (7)	0.90 (2)	1.05 (6)	0.87 (1)
8	1.01 (4)	1.05 (6)	1.21 (8)	0.87 (2)	1.07 (7)	0.94 (3)	1.01 (5)	0.85 (1)
12	1.06 (6)	1.10 (7)	1.35 (8)	0.67 (1)	1.04 (5)	0.95 (3)	0.95 (4)	0.86 (2)
Norway								
1	0.93 (1)	1.02 (6)	0.94 (2)	1.08 (8)	1.02 (5)	0.95 (3)	1.05 (7)	1.00 (4)
4	1.00 (4)	0.93 (1)	1.04 (8)	0.97 (2)	0.98 (3)	1.01 (5)	1.03 (7)	1.03 (6)
8	1.03 (5)	0.92 (2)	1.17 (8)	0.85 (1)	0.98 (3)	1.04 (7)	0.99 (4)	1.04 (6)
12	0.99 (4)	0.91 (1)	1.22 (8)	0.91 (2)	1.02 (7)	1.01 (6)	0.93 (3)	1.01 (5)
Portugal								
1	0.93 (2)	0.94 (3)	0.97 (5)	1.13 (8)	1.09 (7)	0.97 (4)	1.09 (6)	0.89 (1)
4	0.98 (5)	0.97 (3)	1.15 (8)	1.01 (6)	0.96 (2)	1.02 (7)	0.98 (4)	0.93 (1)
8	0.98 (4)	0.99 (5)	1.28 (8)	0.96 (3)	0.91 (2)	1.01 (7)	0.89 (1)	0.99 (6)
12	0.95 (4)	0.95 (5)	1.41 (8)	0.93 (3)	0.85 (2)	0.97 (6)	0.80 (1)	1.14 (7)
Spain								
1	0.90 (2)	0.88 (1)	0.91 (3)	1.06 (7)	1.06 (6)	1.04 (4)	1.10 (8)	1.05 (5)
4	1.10 (7)	1.06 (4)	1.07 (5)	0.69 (1)	0.95 (2)	1.08 (6)	0.96 (3)	1.10 (8)
8	1.20 (8)	1.17 (7)	1.16 (6)	0.49 (1)	0.89 (3)	1.13 (5)	0.84 (2)	1.12 (4)
12	1.18 (5)	1.19 (6)	1.22 (8)	0.53 (1)	0.83 (3)	1.20 (7)	0.74 (2)	1.11 (4)
Switzerland								
1	0.97 (3)	1.05 (8)	0.96 (2)	1.04 (7)	1.03 (6)	0.95 (1)	1.00 (4)	1.01 (5)
4	0.96 (2)	1.02 (6)	1.07 (8)	0.98 (3)	0.99 (5)	0.98 (4)	0.92 (1)	1.07 (7)
8	0.97 (3)	1.08 (7)	1.20 (8)	0.91 (2)	0.98 (5)	0.97 (4)	0.84 (1)	1.04 (6)
12	1.00 (5)	1.07 (6)	1.25 (8)	0.96 (4)	0.85 (2)	0.88 (3)	0.75 (1)	1.23 (7)
United Kingdom								
1	1.19 (8)	0.89 (2)	0.83 (1)	0.95 (4)	1.03 (5)	1.14 (7)	1.06 (6)	0.93 (3)
4	1.00 (5)	0.99 (4)	1.13 (8)	1.02 (7)	0.93 (1)	0.97 (3)	0.96 (2)	1.00 (6)
8	0.98 (5)	0.93 (2)	1.27 (8)	0.99 (7)	0.95 (3)	0.98 (6)	0.98 (4)	0.91 (1)
12	0.95 (3)	0.88 (1)	1.42 (8)	0.97 (5)	0.95 (4)	0.97 (6)	0.99 (7)	0.89 (2)
United States								
1	1.01 (6)	1.00 (5)	0.85 (1)	1.13 (8)	0.96 (2)	1.04 (7)	1.00 (4)	1.00 (3)
4	1.02 (5)	1.03 (7)	1.00 (4)	0.99 (3)	1.08 (8)	0.88 (1)	1.03 (6)	0.97 (2)
8	1.00 (5)	1.04 (6)	1.06 (7)	0.95 (3)	1.14 (8)	0.90 (1)	0.94 (2)	0.97 (4)
12	0.91 (2)	0.92 (3)	1.23 (8)	0.92 (4)	1.21 (7)	0.86 (1)	0.94 (5)	1.01 (6)

The Table contains ratios of RMSPE of the different models to the average RMSPE at horizons $h = 1, 4, 8$ and 12, with ranks of the models in brackets. "AR" is the linear model with constant deterministic seasonality (1); "SUR" is the linear model with seasonal unit roots (2); "SARIMA" is the SARIMA model (3); "PAR" is the periodic model based on (4); "TV-AR" is the linear model with smoothly changing deterministic seasonality (7); "STAR" is the non-linear model with constant deterministic seasonality (9); "TV-STAR" is the nonlinear model with smoothly changing deterministic seasonality (10); "SUR-STAR" is the nonlinear model with seasonal unit roots (11).

Table 2: Summary of Out-of-sample Forecasting Performance

Horizon	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
					Model			
Mean ranks across countries								
1	3.76	3.65	2.29	6.76	4.94	5.06	5.24	4.29
4	4.12	4.00	6.35	3.53	3.71	4.35	4.88	5.06
8	4.53	4.47	7.12	3.06	3.59	5.41	3.24	4.59
12	4.53	4.18	7.41	3.06	3.12	5.29	3.24	5.18
Mean RMSPE across countries								
1	0.97	0.94	0.91	1.10	1.02	1.01	1.02	1.00
4	0.99	0.99	1.06	0.95	0.98	1.00	1.01	1.02
8	1.01	1.00	1.15	0.89	0.95	1.02	0.94	1.01
12	1.00	0.98	1.22	0.87	0.90	1.03	0.89	1.03

The Table contains ratios of MSPE of the different models to the MSPE of an unweighted average forecast at horizon $h = 1$ in individual quarters, with ranks of the models in brackets.

Table 3: Forecast Evaluation: Diebold-Mariano Tests

				Model j				
Model i	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
$h=1$								
AR		0	0	4	3	2	3	3
SUR	2		0	4	4	4	4	3
SARIMA	4	3		9	7	5	5	5
PAR	1	0	0		0	1	0	0
TV-AR	1	0	0	6		1	3	0
STAR	0	0	0	3	2		0	1
TV-STAR	0	0	0	3	1	2		0
SUR-STAR	3	0	0	2	1	3	0	
$h=4$								
AR		1	2	1	0	4	1	3
SUR	1		3	1	0	4	1	3
SARIMA	0	0		0	0	0	0	0
PAR	2	2	3		1	4	1	3
TV-AR	1	0	4	0		1	2	2
STAR	2	2	3	2	1		1	3
TV-STAR	0	0	3	0	1	0		2
SUR-STAR	0	2	2	1	0	1	0	
$h=8$								
AR		1	3	2	0	3	0	2
SUR	4		2	2	0	5	0	4
SARIMA	0	0		0	0	0	0	0
PAR	5	2	9		2	4	1	3
TV-AR	3	2	8	0		4	3	3
STAR	3	1	3	0	0		0	1
TV-STAR	3	2	8	0	2	2		3
SUR-STAR	1	1	2	2	1	3	1	
$h=12$								
AR		3	5	1	0	6	0	5
SUR	4		6	2	1	6	1	7
SARIMA	1	0		0	0	1	1	0
PAR	5	4	9		3	6	0	4
TV-AR	4	4	10	1		6	3	7
STAR	2	2	9	1	0		0	2
TV-STAR	4	3	10	2	6	5		6
SUR-STAR	1	0	5	1	1	4	0	

The table presents pair-wise model comparisons based on out-of-sample forecasting results for the 17 industrial production series and 40 estimation windows. The (i,j)-th entry in the different panels is the number of series for which the Diebold-Mariano statistic rejects the null hypothesis that Model i's and Model j's forecast performance at horizon h as measured by MSPE are equal in favor of the alternative that Model i's performance is superior at the (one-sided) 5% significance level. See Table 1 for model definitions.

Table 4: Forecast Evaluation: Forecast Encompassing Tests

Model i	Model j							
	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
$\underline{h=1}$								
AR		8	6	8	9	14	10	8
SUR	11		7	10	11	9	10	13
SARIMA	13	13		15	15	12	14	15
PAR	2	0	1		2	1	3	0
TV-AR	5	3	2	12		5	15	2
STAR	8	5	3	5	6		4	7
TV-STAR	6	3	0	7	12	5		2
SUR-STAR	4	9	1	4	3	5	4	
$\underline{h=4}$								
AR		13	15	11	12	15	14	14
SUR	12		13	10	11	11	11	15
SARIMA	10	10		6	9	9	10	12
PAR	13	11	15		14	13	14	13
TV-AR	16	14	14	10		13	13	14
STAR	11	7	14	8	10		10	12
TV-STAR	12	10	13	8	11	14		11
SUR-STAR	12	11	14	11	9	10	12	
$\underline{h=8}$								
AR		10	10	8	11	12	12	13
SUR	10		12	7	9	12	11	13
SARIMA	8	8		3	4	8	4	7
PAR	11	11	13		10	12	11	12
TV-AR	14	14	14	10		12	12	14
STAR	11	8	11	6	8		12	8
TV-STAR	14	13	14	10	10	14		13
SUR-STAR	11	7	10	10	6	12	7	
$\underline{h=12}$								
AR		7	8	3	7	8	8	11
SUR	10		11	6	7	9	8	11
SARIMA	5	4		3	2	5	3	4
PAR	7	7	9		8	10	7	10
TV-AR	14	12	11	10		13	7	13
STAR	5	3	8	4	6		8	8
TV-STAR	14	13	11	10	11	15		15
SUR-STAR	7	5	9	4	6	12	5	

The table presents pair-wise model comparisons based on out-of-sample forecasting results for the 17 industrial production series and 40 estimation windows. The (i,j)-th entry in the different panels is the number of series for which the Diebold-Mariano type forecasting encompassing statistic of Harvey, Leybourne and Newbold (1998) does not reject the null hypothesis that Model i's forecast at horizon h encompasses Model j's forecast at the 5% significance level. See Table 1 for model definitions.

Table 5: Out-of-sample Forecasting Performance per Quarter

Quarter	AR	SUR	SARIMA	PAR	TV-AR	STAR	TV-STAR	SUR-STAR
Canada								
1	0.91 (3)	0.87 (2)	0.86 (1)	1.05 (5)	1.43 (8)	1.19 (6)	1.36 (7)	0.96 (4)
2	1.04 (2)	1.23 (6)	1.07 (3)	1.65 (8)	1.14 (5)	1.13 (4)	0.85 (1)	1.51 (7)
3	0.82 (1)	0.84 (2)	0.95 (4)	1.67 (8)	1.35 (6)	0.94 (3)	1.35 (7)	1.16 (5)
4	0.98 (1)	1.01 (2)	1.03 (4)	1.12 (6)	1.35 (7)	1.07 (5)	1.43 (8)	1.03 (3)
France								
1	1.38 (8)	0.96 (2)	1.03 (5)	0.90 (1)	0.99 (4)	1.15 (7)	0.97 (3)	1.11 (6)
2	1.36 (7)	0.97 (5)	0.91 (1)	1.08 (6)	0.97 (4)	1.43 (8)	0.93 (2)	0.97 (3)
3	0.95 (2)	0.95 (3)	0.94 (1)	1.13 (6)	1.21 (7)	1.04 (4)	1.29 (8)	1.08 (5)
4	1.47 (5)	0.84 (1)	0.90 (2)	1.20 (3)	1.50 (6)	1.64 (7)	1.71 (8)	1.22 (4)
Germany								
1	0.92 (1)	0.96 (3)	0.96 (4)	1.07 (5)	1.32 (6)	0.93 (2)	1.32 (7)	1.32 (8)
2	1.20 (7)	1.28 (8)	1.08 (5)	1.03 (4)	0.90 (2)	0.83 (1)	0.91 (3)	1.12 (6)
3	1.09 (2)	0.75 (1)	1.16 (5)	2.24 (8)	1.37 (6)	1.13 (4)	1.39 (7)	1.10 (3)
4	1.31 (5)	1.06 (3)	1.09 (4)	1.42 (6)	1.01 (2)	1.87 (8)	0.97 (1)	1.70 (7)
Italy								
1	0.82 (1)	1.04 (4)	0.99 (3)	1.60 (8)	1.22 (7)	0.93 (2)	1.10 (5)	1.13 (6)
2	1.14 (5)	1.22 (7)	0.89 (2)	1.11 (3)	1.12 (4)	1.15 (6)	0.89 (1)	1.38 (8)
3	1.09 (5)	1.10 (6)	0.98 (3)	0.96 (2)	0.93 (1)	1.10 (7)	1.01 (4)	1.19 (8)
4	1.05 (2)	1.26 (7)	1.12 (4)	1.48 (8)	1.16 (6)	1.03 (1)	1.07 (3)	1.16 (5)
Japan								
1	1.01 (3)	1.24 (8)	1.21 (7)	1.09 (4)	0.95 (1)	1.16 (5)	0.98 (2)	1.20 (6)
2	0.88 (2)	1.05 (5)	1.06 (6)	1.01 (3)	1.02 (4)	1.55 (8)	1.38 (7)	0.84 (1)
3	1.40 (6)	0.89 (3)	0.74 (1)	1.07 (5)	1.50 (8)	1.49 (7)	1.06 (4)	0.87 (2)
4	1.06 (4)	1.00 (1)	1.04 (3)	1.50 (7)	1.04 (2)	1.59 (8)	1.12 (6)	1.07 (5)
United Kingdom								
1	1.11 (5)	0.78 (1)	1.00 (3)	1.10 (4)	1.13 (6)	1.23 (8)	1.16 (7)	0.92 (2)
2	1.19 (6)	0.89 (1)	0.95 (2)	1.08 (5)	1.32 (7)	0.98 (3)	1.35 (8)	1.06 (4)
3	1.28 (7)	1.28 (8)	0.95 (1)	1.18 (3)	1.23 (5)	1.22 (4)	1.27 (6)	1.13 (2)
4	2.22 (8)	1.19 (5)	1.07 (2)	1.14 (4)	1.06 (1)	2.20 (7)	1.09 (3)	1.29 (6)
United States								
1	1.32 (8)	1.22 (6)	1.06 (3)	1.14 (4)	0.98 (1)	1.32 (7)	1.05 (2)	1.21 (5)
2	1.23 (6)	1.24 (7)	1.22 (5)	1.30 (8)	1.15 (4)	1.13 (3)	0.99 (2)	0.87 (1)
3	0.98 (2)	1.00 (3)	0.84 (1)	1.35 (7)	1.21 (6)	1.07 (4)	1.49 (8)	1.17 (5)
4	1.19 (4)	1.24 (5)	0.84 (1)	1.46 (8)	1.13 (3)	1.33 (6)	0.98 (2)	1.35 (7)

The Table contains ratios of MSPE of the different models to the MSPE of an unweighted average forecast at horizon $h = 1$ in individual quarters, with ranks of the models in brackets.