Journal of Econometrics 55 (1993) 173–200. North-Holland

Forecasting time series with common seasonal patterns*

Fabio Canova

Brown University, Providence, RI 02912, USA

European University Institute, I-50010 San Domenico di Fiesole, Italy

This paper proposes a methodology for modelling and forecasting series which possess common patterns at seasonal and/or other frequencies. The approach is in the Bayesian autoregression tradition originally developed by Litterman (1980), Doan, Litterman, and Sims (1984), and Sims (1989), and builds common patterns directly into the prior of the coefficients of the model by means of a set of uncertain linear restrictions. To gauge the usefulness of the approach, the procedure is applied to the problem of forecasting a small vector of national industrial production indices.

1. Introduction

The observations that many economic time series exhibit pronounced trends and seasonals and that some also show common features across time both in the short and in the long run were recognized long ago [e.g., Burns and Mitchell (1946)]. However, it is only recently that time series econometricians have formalized in econometric models the concept of common comovements at particular frequencies and the idea that common factors may affect the behavior of several economic aggregates. For example, Engle and Granger (1987) and Stock and Watson (1988) have examined common long-run patterns in integrated series, Hylleberg, Engle, Granger, and Yoo (1990) common patterns at seasonal frequencies for time series which are seasonally integrated, Kang (1988) common deterministic factors in trend-stationary series, and Gourieroux and Paucelle (1989) common long-run patterns in stationary series.

*This paper was written for the conference on 'Seasonality and Econometric Models', May 11–12, 1990, Montreal, and was also presented at the 1990 Meetings of the Society for Economic Dynamic and Control, Minneapolis. I would like to thank Tim Erickson for valuable conversations, Eric Ghysels (the editor), Adrian Pagan, Chris Sims, Arnold Zellner, and two anonymous referees for useful suggestions. The comments and the suggestions of John Geweke were instrumental in eliminating errors from an earlier draft and in improving the focus of the paper.

This paper incorporates the concept of common patterns at seasonal and at other frequencies into a Bayesian forecasting model of the Doan, Litterman, and Sims (DLS) (1984) variety. There are two reasons to do this. First, a researcher's beliefs concerning the existence of common patterns at certain frequencies can be embodied quite naturally in a prior for the coefficients of the model, and their support in the data is easily summarized in a posterior. Second, since DLS-type models are well-known to provide good forecasts in the short run [see Litterman (1986) and McNees (1986) for seasonally adjusted macro data, Canova (1992) and Raynauld and Simonato (1993) for seasonally unadjusted macro data], it is of interest to see whether accounting for common patterns at certain frequencies improves the ability of these models to forecast at seasonal and long horizons.[1]

Canova (1992) demonstrated that beliefs concerning the existence of peaks in the spectral density at seasonal frequencies imply a set of uncertain linear restrictions on the coefficients of the AR representation of univariate time series. This information was incorporated into the prior for the coefficients of the model by means of Theil's mixed type estimation. This paper extends that analysis by characterizing the restrictions implied by beliefs concerning the existence of common patterns at seasonal and nonseasonal frequencies on the AR representation of a panel of time series and discusses the incorporation of these restrictions into the prior for the coefficients of the model [Raynauld and Simonato (1993) provide alternative ways to account for seasonals in a Bayesian VAR model]. The focus is on a panel of same time series because the idea of common factors quite naturally applies to this framework of analysis [see Thisted and Wecker (1981) and Zellner and Hong (1989) for a similar approach]. However, the methodology is general and well-suited to handle 'random field' data [of the type employed by Quah (1989)] or standard VAR models.

The modelling approach employed to characterize the presence of common patterns at various frequencies is advantageous in several respects. First, it does not impose restrictions on the way seasonals and other components of a series interact [see Ghysels (1988) and Hansen and Sargent (1990) for criticisms of the restrictions imposed by traditional orthogonal decompositions of time series from the point of view of dynamic economic theory]. Second, it does not constrain the log spectrum of the estimated seasonal component to have the same power at each seasonal frequency [see Sims (1979) for this feature of ARIMA modelling of seasonality]. In addition, since beliefs concerning the existence of common patterns are modelled probabilistically, the data is left to decide the relative support for each type of restriction [see Drèze (1975) for a similar approach in identifying simultane-

[1]Canova (1992) reports, for example, that a model with a simple linear time trend plus seasonal dummies outperforms a univariate version of the DLS model in forecasting at twelve-quarter horizons over the period 1969.4–1981.4 for housing start, consumer durables, fixed investments, business inventories, and final sales in the U.S.

ous equation models]. Finally, the methodology is flexible enough to capture general forms of common seasonality and can endogenously account for their evolution over time. This avoids the use of judgmental procedures or extensive model respecification to 'repair' relationships which break down from time to time.

I judge the usefulness of common pattern restrictions using out-of-sample forecasting criteria. For this purpose I will be interested in making posterior probability statements concerning functions of the parameters of the model. For any function $f(\theta)$ of the parameters θ, such statements require the construction of objects of the form

$$\mathrm{E}(f(\theta)|\mathscr{A}, y, x) = \int_{\mathscr{A}} f(\theta) P(\theta|x, y)\,\mathrm{d}\theta, \tag{1}$$

where $\mathscr{A}$ is a subset of the parameter space Θ, $P(\theta|x, y)$ is the posterior of the θ given data on x_t and y_t, where x_t are the observables at t and y_t are past values of the observables. For the model of this paper the posterior of θ has the form

$$P(\theta|x_t, y_t) = \frac{\int \mathscr{G}(x_t|y_t, \beta_t)\mathscr{F}(\beta_t|\theta)\mathscr{Q}(\theta)\,\mathrm{d}\beta_t}{\int\int \mathscr{G}(x_t|y_t, \beta_t)\mathscr{F}(\beta_t|\theta)\mathscr{Q}(\theta)\,\mathrm{d}\beta_t\,\mathrm{d}\theta}$$

$$\propto \mathscr{L}(\theta|x_t, y_t)\mathscr{Q}(\theta), \tag{2}$$

where $\mathscr{L} \propto \int \mathscr{G}(x_t|y_t, \beta_t)\mathscr{F}(\beta_t|\theta)\mathrm{d}\beta_t$ is the marginalized likelihood of θ given data on x and y and $\mathscr{Q}(\theta)$ is the prior density on the parameters of the model. $\mathscr{G}$ is the statistical model linking the observables to a vector of time-varying coefficients (the β's) and to past values of the observables. Prior information about the relationship between the vector of time-varying β_t and a set of unknown but time-invariant parameters θ is expressed via the conditional density $\mathscr{F}(\beta_t|\theta)$. Although I employ a hierarchical structure with only one intermediate layer of uncertainty, specifications which include several dimensions of uncertainty are possible [see Monahan (1983)].

In the particular framework used here little is known about the properties of the θ's except, perhaps, the range of the support. Therefore the analysis will proceed by assuming that $\mathscr{Q}(\theta)$ is flat in some hypercube. In this situation the posterior density of the parameters is proportional to the marginalized likelihood function [Zellner (1971)] and inference can proceed by directly integrating over the region of interest of the likelihood function. Also, the marginalized likelihood function can be used as a diagnostic measure of fit between the remaining component of the prior $\mathscr{F}(\beta_t|\theta)$ and the data [see Doan, Litterman, and Sims (1984)].

The analysis is thus conducted in two steps: first, several parameter vectors, all having equal prior probability, are passed through a given data set and the coefficients of the model are updated recursively using the Kalman filter. For each parameter vector the predictions generated with recursively updated coefficients are compared with the actual data, and the magnitude of the resulting forecast error is used to gauge the relative support for the choice of parameters [the 'fit' of $\mathscr{F}(\beta_t|\theta)$ to the data]. As part of this routine, a sensitivity analysis is employed to assess how the measure of fit changes when certain parameters are fixed at their default values rather than at their optimal levels or when the functional specification of the prior is altered [see Carlin, Dempster, and Jonas (1985) for related procedure]. As a by-product of the analysis, a numerical approximation of the marginal posterior density of the parameters is obtained. Second, choosing $\mathscr{A}$ to be the region of the parameter space in a specified neighborhood of the peak of the likelihood, I construct confidence regions around the mean of out-of-sample forecasting statistics using a Monte Carlo approximation to the integral eq. (1).

The rest of the paper is organized as follows: section 2 describes the statistical model and the structure of $\mathscr{F}(\beta_t|\theta)$. Section 3 derives the restrictions implied by the presence of common patterns at seasonal frequencies on the AR coefficients of the model and discusses how to combine this source of information with other available prior information on the coefficients. A sketch of the restrictions implied by beliefs concerning the existence of common patterns at other frequencies is also provided. Section 4 discusses inference. Section 5 provides an example and assesses the contribution of various features of the prior to the forecasting performance of the model. Section 6 contains the conclusions.

2. The statistical model and the DLS-type prior

The statistical model $\mathscr{S}(x_t|y_t, \beta_t)$ is given by

$$x_t = a_t(l)x_{t-1} + c_t D_t + u_t, \tag{3}$$

where $\beta_t' = [\beta_{1t}, \ldots, \beta_{Nt}]$ with $\beta_{nt} = [a_{n11t}, a_{n12t}, \ldots, a_{n1Kt}, \ldots, a_{nN1t}, a_{nN2t}, \ldots, a_{nNKt}, c_{n1t}, \ldots, c_{nSt}]$; $y_t = I \times Y_t$ with $Y_t = [x_{1t-1}, x_{1t-2}, \ldots, x_{1t-K}, \ldots, x_{Nt-1}, x_{Nt-2}, \ldots, x_{Nt-K}, D_{1t}, \ldots, D_{St}]$; I is a $N \times N$ identity matrix where N is the dimension of the vector of time series, $1 \leq n \leq N$; l is the lag operator; K is the maximum number of lags allowed for each n, $1 \leq k \leq K$; T is the number of observations, $1 \leq t \leq T$; and S is the number of seasons in the year, $1 \leq s \leq S$. Here $u_t = [u_{1t}, u_{2t}, \ldots, u_{Nt}]$ is a vector of serially uncorrelated, conditionally Gaussian disturbances with zero mean and covariance matrix Σ_u (typical element $\sigma_{nn'}$) and c_t are time-varying coefficients on the seasonal dummies D_t [see, e.g., Hylleberg (1986) for a justification of a time-varying coefficients approach to model seasonality].

Prior information about the law of motion of the coefficients and their relationship with the vector of hyperparameters θ given by

$$\beta_t = (\theta_0 \times I)\beta_{t-1} + ((1-\theta_0) \times I)\beta_0 + \varepsilon_t, \tag{4}$$

where I is a $N(NK+S) \times N(NK+S)$ identity matrix and θ_0 regulates the decay of the coefficients. For $\theta_0 < 1$, the prior law of motion of the coefficients in each equation shrinks the previous period's value toward the information available at time zero. I assume that $\beta_0 \sim \mathcal{N}(\mathrm{E}(\beta_0), \mathrm{var}(\beta_0))$ and that $\varepsilon_t \sim \mathcal{N}(0, \theta_1 \times \mathrm{var}(\beta_0))$. θ_1 controls the amount of randomness injected in the variance of the coefficients at each t. For $\theta_1 = 0$ the coefficients are constant over time.

The mean of the coefficients at time zero is characterized by two typical elements, $\mathrm{E}(a_{nn'k0})$ and $\mathrm{E}(c_{ns0})$, which are assumed to have the following structure:

$$\begin{aligned} \mathrm{E}(a_{nn'k0}) &= \theta_7 \quad \text{if} \quad k=1, \quad n=n', \\ &= \theta_8 \quad \text{if} \quad k=S, \quad n=n', \\ &= 0 \quad \text{otherwise}, \end{aligned} \tag{5}$$

$$\mathrm{E}(c_{ns0}) = 0 \qquad \forall n, s, \tag{6}$$

where θ_7 and θ_8 are parameters regulating the prior means on the first own lagged and on the first own seasonal lagged AR coefficients.

At this stage of the formulation, I assume that the covariance matrix of the coefficients at time zero is diagonal. The next section shows that frequency domain considerations provide prior information about the off-diagonal elements of this matrix. The typical elements of $\mathrm{var}(\beta_0)$ are of two types, $\mathrm{var}(a_{nn'k0})$ and $\mathrm{var}(c_{ns0})$, and are assumed to have the following structure:

$$\mathrm{var}(a_{nn'k0}) = \frac{\theta_2\theta_4\theta_5}{k^{\theta_3}} \times \frac{\sigma_{n'n'}}{\alpha_{n,n}} \quad \text{with} \quad \begin{aligned} &\theta_5 = 1 \quad \text{if} \quad n=n', \\ &\theta_4 \neq 1 \quad \text{if} \quad n=n', \\ &\qquad \text{and} \quad k=hS, \end{aligned} \tag{7}$$

$$\mathrm{var}(c_{ns0}) = \theta_2 \times \theta_6 \qquad s = 1,2,\dots,S, \tag{8}$$

where $h = 1,\dots,[K/S]$. Here θ_2 is a general tightness parameter regulating the concentration of the prior variance of each coefficient around its prior median; θ_3 controls how the prior distribution of each AR coefficient becomes concentrated around its median as the lag length k increases; θ_4 controls the concentration of the prior distribution around each median of own seasonal AR coefficients in each equation; θ_5 controls the concentration of the prior distribution around each median of lagged AR coefficients of other variables in each equation; and θ_6 is the tightness on the prior variance on the coefficients of each seasonal dummy. Note that when $\theta_5 = 0$, the elements of x_0 are *a priori* restricted to be only contemporaneously related. Finally, $\sigma_{n'n'}/\sigma_{n,n}$ is a scaling factor for the variance of the AR coefficients of other variables in each question. These scaling factors will be estimated from the data to tune up the prior to the particular application.

Therefore, there are nine parameters regulating the evolution of the coefficients of a $N \times 1$ vector of time series.

Eqs. (5)–(8) build a lot of symmetry into the model. Information about the coefficients at time zero is assumed to be the same for each equation (apart from a scaling factor in the variance of the AR coefficients). This restriction is justified by the observation that for a panel of time series one *a priori* expects the coefficients and their variances to be similar [see Thisted and Wecker (1981) for a similar shrinkage approach]. Note also that this formulation is useful for our purposes because it imposes a common pattern restriction on the coefficients of the seasonal dummies in each equation by means of an exchangable-type prior [see Lindley and Smith (1972) and Zellner and Hong (1989) for a justification of this type of prior].

The specification chosen for $\mathscr{F}(\beta_t|\theta)$ differs in three respects from the one employed by Doan, Litterman, and Sims (1984) and Sims (1989). First, the prior mean on the first own seasonal lag coefficient in each equation is allowed to be different from zero and $\theta_7 + \theta_8$ is not restricted to be equal to one. Second, the prior variance on the own seasonal lagged AR coefficients is separately parameterized. Third, deterministic seasonal dummies are included in the statistical model and the evolution of their coefficients is treated in the same way as the evolution of the AR coefficients. Therefore, the DLS prior can be recovered from this framework by setting $\theta_7 = 1$, $\theta_8 = 0$, $\theta_4 = 1$, $\forall k$, and $\theta_6 = 0$, $\forall s \neq 1$. Note also that while the dummies are intended to capture deterministic or very slowly evolving seasonals, the restrictions considered in the next section are designed to account for remaining stochastic forms of seasonality.

Finally, letting $\rho_t = -\varepsilon_t$, $\Sigma_\rho = \mathrm{E}_t[\rho_t \rho_t']$, $R = I$, and $A_t = (\theta_0 \times I)\beta_{t-1} + ((1-\theta_0) \times I)\beta_0$, it is easy to see that (4) can be written as

$$A_t = R\beta_t + \rho_t, \qquad \rho_t \sim \mathscr{N}(0, \Sigma_\rho), \tag{9}$$

which shows that our prior specification on the β's comes in the form of a set of uncertain linear restrictions on the coefficients of the model at each t.

3. The restrictions implied by the presence of common patterns at seasonal frequencies

Before characterizing the restrictions on the coefficients of model (3) implied by the presence of common stochastic patterns at seasonal frequencies, a definition of common stochastic seasonality is needed. For this purpose let $F_{n,n}(\omega)$ and $F_{n',n'}(\omega)$ be the (pseudo)spectral densities of $x_{n,t}$ and $x_{n',t}$, respectively. Let $F_{n,n'}(\omega)$ be the (pseudo)cross-spectral density of $x_{n,t}$ and $x_{n',t}$ and let ω_s be the seasonal frequencies, $s = 1, 2, \ldots, S/2$. Finally, let $Co_{n,n'}(\omega) = [F_{n,n'}(\omega)]^2/F_{n,n}(\omega)F_{n'n'}(\omega)$ be the coherence between x_{nt} and $x_{n't}$ at frequency ω. Following Granger (1979) and Granger and Weiss (1983), we will say that x_{nt} and $x_{n't}$ possess common patterns at some ω_s, if $F_{n,n}(\omega)$ and $F_{n'n'}(\omega)$ have a peak (or a large mass) at ω_s and if $Co_{n,n'}(\omega_s) \approx 1$. Therefore, the vector x_t possesses common patterns at some seasonal frequencies if the real part of the (pseudo)spectral density matrix has peaks in all its elements at these frequencies.

Next, I show that the existence of peaks in the (pseudo)spectral density matrix at seasonal frequencies implies restrictions on the coefficients of the AR representation of the vector x_t. Let $x_t = \mu(l)v_t + \hat{x}$ be the MA representation for x_t, where v_t is a serially uncorrelated random noise with $\text{cov}(v_t, v_{t'}) = \Sigma_v$ if $t = t'$ and zero otherwise, $\mu(l)$ is a $N \times N$ matrix polynomial in the lag operator of order $Q \leq \infty$, and $\hat{x}$ is the singular component of x_t. Tedious but straightforward algebra indicates that the (n, n') elements of the (pseudo)spectral density matrix $F(\omega)$ can be written as

$$F_{n,n'}(\omega) = \sum_{j=1}^{N} \sum_{l=1}^{N} \sigma_{jl} \left[\sum_{q=1}^{Q} \mu_{njq}\mu_{n'lq} + \sum_{h=1}^{Q-1} \mathrm{e}^{-i\omega h} \left(\sum_{q=1}^{Q-h} \mu_{njq}\mu_{n'l(q+h)} \right) \right.$$
$$\left. + \sum_{h=1}^{Q-1} \mathrm{e}^{i\omega h} \left(\sum_{q=1}^{Q-h} \mu_{nj(q+h)}\mu_{n'lq} \right) \right]$$
$$+ \sigma_{nn'} + \sum_{j=1}^{N} \sum_{q=1}^{Q} \left(\mathrm{e}^{i\omega q}\sigma_{nj}\mu_{n'jq} + \mathrm{e}^{-i\omega q}\sigma_{jn'}\mu_{njq} \right). \tag{10}$$

For the sake of presentation let $S = 4$ so that $\pi/2$ and π are the seasonal frequencies. Using (10) for $n' = n$ and neglecting nonlinear terms, a peak in $F_{n,n}(\omega)$ at $\pi/2$ requires that for every n

$$\sum_{j=1}^{N} \sum_{q=1}^{[Q/2]} (-1)^q \mu_{nj(2q)}\sigma_{nj} \geq M_1, \tag{11}$$

while a peak at π requires

$$\sum_{j=1}^{N} \sum_{q=1}^{Q} (-1)^q \mu_{njq}\sigma_{nj} \geq M_2, \tag{12}$$

where M_1 and M_2 are large constants. If $Co_{n,n'}(\omega_s)$ is close to one, and neglecting nonlinear terms, it must also be the case that

$$\sum_{j=1}^{N}\sum_{q=1}^{[Q/2]}(-1)^{q}\left[\mu_{nj(2q)}H_{1j}+\mu_{n'j(2q)}H_{2j}\right]=-\tfrac{1}{2},$$

and/or

$$\sum_{j=1}^{N}\sum_{q=1}^{Q}(-1)^{q}\left[\mu_{njq}H_{1j}+\mu_{n'jq}H_{2j}\right]=-\tfrac{1}{2}, \tag{13}$$

where the first restriction holds at $\pi/2$ and the second at π, and where

$$H_{1j}=\left(\frac{\sigma_{n,n'}\sigma_{j,n'}-\sigma_{n',n'}\sigma_{n,j}}{\sigma_{n,n'}^{2}-\sigma_{n,n}\sigma_{n',n'}}\right),\qquad H_{2j}=\left(\frac{\sigma_{n,n'}\sigma_{n,j}-\sigma_{n,n}\sigma_{n',j}}{\sigma_{n,n'}^{2}-\sigma_{n,n}\sigma_{n',n'}}\right).$$

If there exists a real valued constant $\kappa\geq 1$ such that $\kappa^{-1}\times I\leq\Re[F(\omega)]\leq\kappa\times I$ for almost all ω, where I is a $N\times N$ identity matrix and where $A\leq B$ means that $B-A$ is positive semidefinite, then an AR representation for x_t exists [see Rozanov (1967, p. 77)]. Moreover, the presence of common patterns at seasonal frequencies in x_{nt} and $x_{n't}$ implies that

$$\sum_{j=1}^{N}\sum_{g=0}^{\infty}(-1)^{g}\eta_{nj(2g)}\frac{\sigma_{nj}}{\sigma_{n,n}}\approx 0, \tag{14}$$

$$\sum_{j=1}^{N}\sum_{m=0}^{\infty}(-1)^{m}\eta_{n'j(2m)}\frac{\sigma_{n'j}}{\sigma_{n'n'}}\approx 0, \tag{15}$$

$$\sum_{j=1}^{N}\left[\sum_{m=0}^{\infty}(-1)^{m}\eta_{n'j(2m)}H_{2j}+\sum_{g=0}^{\infty}(-1)^{g}\eta_{nj(2g)}H_{1j}\right]\approx 0, \tag{16}$$

$$\sum_{j=1}^{N}\sum_{g=0}^{\infty}(-1)^{g}\eta_{njg}\frac{\sigma_{nj}}{\sigma_{nn}}\approx 0, \tag{17}$$

$$\sum_{j=1}^{N}\sum_{m=0}^{\infty}(-1)^{m}\eta_{n'jm}\frac{\sigma_{n'j}}{\sigma_{n'n'}}\approx 0, \tag{18}$$

$$\sum_{j=1}^{N}\left[\sum_{m=0}^{\infty}(-1)^{m}\eta_{n'jm}H_{2j}+\sum_{g=0}^{\infty}(-1)^{g}H_{1j}\eta_{njg}\right]\approx 0, \tag{19}$$

where $\eta(l)=\mu(l)^{-1}$, (14)–(16) hold at $\omega=\pi/2$, and (17)–(19) hold at $\omega=\pi$.

To gain some intuitive understanding of the content of these restrictions consider the case where $\mu(l)$ is diagonal so that only contemporaneous effects among the components of x_t are allowed, and concentrate attention on the restrictions emerging at $\pi/2$. In this case a peak in the spectral density of x_{nt} implies

$$\sum_{g=1}^{\infty} (-1)^g \eta_{nn(2g)} \approx -1, \tag{20}$$

while the coherence of x_{nt} and $x_{n't}$ being close to one implies

$$\sum_{m=1}^{\infty} (-1)^m \eta_{n'n'(2m)} + \sum_{g=1}^{\infty} (-1)^g \eta_{nn(2g)} \approx -2. \tag{21}$$

Therefore a peak in the spectral density of x_{nt} ($x_{n't}$) at $\pi/2$ requires that a particular linear combination of its own lagged AR coefficients is close to minus one, while the coherence of x_{nt} and $x_{n't}$ is close to one if a linear combination of the own lagged coefficients of the two variables is close to minus two.[2] To put it in another way, the presence of peaks in the spectral density and of high coherence at seasonal frequencies implies that a linear combination of the seasonal AR coefficients is small. However, as one of the referees pointed out, this intuition does not carry through to the general case. When $\mu(l)$ is not diagonal and the coherence is high, peaks in the spectral density at seasonal frequencies do not imply dips in the Fourier transform of the AR operator on own lags at seasonal frequencies. Instead, as eq. (13) indicates, peaks in the spectral density at seasonal frequencies require that a linear combination of the seasonal AR coefficients on *all* the variables in each equation should be small.

I treat (14)–(19) as probabilistic constraints for two reasons. First, even for series which are known to be seasonal and to move together seasonally, it is not known *a priori* at which frequency the restrictions apply. Second, even when the restrictions are applicable to a particular frequency, the size of the peaks and the closeness of the coherence to one are series dependent. By treating the restrictions as stochastic, we introduce flexibility in the specification and allow the data to depart from (14)–(19) if the information contained

[2]The coherence restriction can be satisfied even though no peaks in the spectral density of x_{nt} and $x_{n't}$ appear at that frequency, provided that $F_{n,n}(\omega)$ and $F_{n'n'}(\omega)$ are similar.

in the restrictions is not pertinent to the pair of time series under consideration.

The random term in each constraint is modelled as a mean zero stochastic variable whose variance represents the researcher's prior confidence on the exactness of that particular restriction. For example, if one *a priori* believes that common patterns appear at only one seasonal frequency, one can capture this idea by imposing a large variance on the restrictions at other seasonal frequencies. Also, if one *a priori* believes that common patterns at seasonal frequencies are of an evolutive nature, one expects the real part of $[F(\omega)]$ to display an even distribution of power over seasonal bands. A relatively large variance on each of the restrictions may capture this belief. Finally, although the modelling approach discussed in this paper is ill-suited to handle unit root processes, we can also capture *a priori* beliefs regarding seasonal cointegration [Hylleberg, Engle, Granger, and Yoo (1990)]. Seasonal cointegration at ω_s occurs if a spike of infinite height exists in the real part of $[F(\omega_s)]$. Therefore, the presence of seasonal cointegration at ω_s can be approximately captured here by selecting the variance of all the restrictions at ω_s to be equal to zero. This is because, if a researcher believes that a vector of time series is seasonally cointegrated at ω_s, there will be no uncertainty regarding the existence and the size of the peak in the spectral density matrix at that frequency.

3.1. Combining the two sets of prior information

To combine the information contained in (4)–(9) with the restrictions derived in the previous subsection into a new prior specification for the coefficients of the model (3) which takes into account the presence of common stochastic patterns at seasonal frequencies, note that (14)–(19) can be written as

$$r = \hat{R}\beta_t + \zeta_t, \qquad \zeta_t \sim \mathcal{N}(0, \Sigma_\zeta), \tag{22}$$

where $\Sigma_\zeta = \theta_9 \times \operatorname{diag}(\sigma^2_{\zeta ii})$ and where the parameter θ_9 represents the general tightness of the restrictions. For quarterly data, r is a $6M \times 1$ vector and $\hat{R}$ is a $6M \times N(NK+4)$ matrix, with M being the number of pairs to which the seasonal restrictions apply, $0 < M < N(N-1)/2$. For example, if common patterns are believed to exist at both seasonal frequencies for $n = 1$ and $n' = 2$, then r is a 6×1 vector with entries $[-1, -1, -2, -1, -1, -2]$ and $\hat{R}$ is a $6 \times N(NK+4)$ matrix with entries shown on the next page:

$$
\left[\begin{array}{cccccccccccccccccccc}
0 & -1 & 0 & 1 & 0 & \dots & 0 & -\sigma_{12} & 0 & \sigma_{12} & 0 & \dots & 0 & -\sigma_{13} & 0 & \sigma_{13} & 0 & \dots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & -H_{13} & 0 & H_{13} & 0 & \dots & 0 & 0 \\
-1 & 1 & -1 & 1 & -1 & \dots & -\sigma_{12} & \sigma_{12} & -\sigma_{12} & \sigma_{12} & -\sigma_{12} & \dots & -\sigma_{13} & \sigma_{13} & -\sigma_{13} & \sigma_{13} & -\sigma_{13} & \dots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
-1 & 1 & -1 & 1 & -1 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & -H_{13} & H_{13} & -H_{13} & H_{13} & -H_{13} & \dots & 0 & 0
\end{array}\right.
$$

$$
\begin{array}{cccccccccccccccccccc}
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
0 & -\sigma_{21} & 0 & \sigma_{21} & 0 & \dots & 0 & -1 & 0 & 1 & 0 & \dots & 0 & -\sigma_{23} & 0 & \sigma_{23} & 0 & \dots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & -1 & 0 & 1 & 0 & \dots & 0 & -H_{23} & 0 & H_{23} & 0 & \dots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\
-\sigma_{21} & \sigma_{21} & -\sigma_{21} & \sigma_{21} & -\sigma_{21} & \dots & -1 & 1 & -1 & 1 & -1 & \dots & -\sigma_{23} & \sigma_{23} & -\sigma_{23} & \sigma_{23} & -\sigma_{23} & \dots & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & -1 & 1 & -1 & 1 & -1 & \dots & -H_{23} & H_{23} & -H_{23} & H_{23} & -H_{23} & \dots & 0 & 0
\end{array}
$$

$$
\left.\begin{array}{cccc}
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0
\end{array}\right]
$$

where the first two blocks of the matrix have dimensions $6 \times (NK+4)$, while the last block has dimension $6 \times (N-2)(NK+4)$.

Letting $z = [A_t, r]$, $Z' = [I, \hat{R}]$, $e_t = [\rho_t, \zeta_t]$, and stacking (9) and (22), we obtain

$$z = Z'\beta_t + e_t. \tag{23}$$

Using the logic of Theil's mixed type of estimation, it is immediate to see that the new $\mathscr{F}(\beta_t | \theta)$ has the form

$$\beta_t \sim \mathscr{N}(\tilde{\beta}_t, \Psi), \tag{24}$$

where

$$\tilde{\beta}_t = \left[\Sigma_\rho^{-1} + \hat{R}'\Sigma_\zeta^{-1}\hat{R}\right]^{-1}\left[\Sigma_\rho^{-1}A_t + \hat{R}'\Sigma_\zeta^{-1}r\right],$$

$$\Psi = \left[\Sigma_\rho^{-1} + \hat{R}'\Sigma_\zeta^{-1}\hat{R}\right]^{-1} = \Sigma_\rho - \Sigma_\rho\hat{R}'\left[\Sigma_\zeta + \hat{R}\Sigma_\rho\hat{R}'\right]^{-1}\hat{R}\Sigma_\rho. \tag{25}$$

The procedure therefore shrinks the original prior mean of the coefficients toward the vector r. In addition, the prior covariance matrix of the coefficients is modified. The restrictions describing the existence of peaks in $F_{n,n}$ and $F_{n'n'}$ at some $\omega = \omega_s$ leave the diagonal elements of Σ_ρ unchanged but introduce nonzero and alternating in sign values in the off-diagonal elements of the diagonal blocks (n,n) and (n',n') of Σ_ρ. On the other hand, the restriction describing how close to one is the coherence of x_{nt} and $x_{n't}$ at ω_s introduces values different than zero in the off-diagonal elements of the (n,n') and (n',n) blocks of Σ_ρ which span seasonal lags.

The ratio θ_2/θ_9, i.e., the ratio of the general tightness of Σ_ρ to the general tightness of Σ_ζ, represents the relative confidence of the two types of information and determines the extent to which Σ_ρ is modified. For θ_2/θ_9 small, the new prior covariance matrix primarily contains information coming from (4)–(8), while for θ_2/θ_9 large, the new prior covariance matrix on the coefficients is tilted toward the information contained in the frequency domain restrictions.

A few features of the approach to modelling common seasonal patterns employed in this paper should be noted. First, the procedure allows both common deterministic and common stochastic seasonal patterns in the prior of the coefficients. Second, unlike multiplicative ARIMA models or transfer function models, this framework does not *a priori* restrict the frequency domain representations of the estimated seasonal patterns to all have the same shape or the same height at each frequency [see Sims (1979)]. Third, the approach does not impose restrictions on the way seasonals interact with other time series components of each x_{nt}. This is an important feature since Ghysels (1988), Hansen and Sargent (1990), and others have pointed out that, in general, dynamic economic models do not result in time series for the endogenous variables where the traditional orthogonal decomposition in trend, seasonal, and irregulars apply. Finally, the methodology is flexible enough to capture several forms of common patterns at seasonal frequencies (almost deterministic patterns, time-varying patterns, very smooth or semi-periodic patterns, as well as common patterns which appear at only one specific seasonal frequency).

3.2. Other frequency domain restrictions

The approach proposed in the previous two subsections can also be used to account for other frequency domain features in the prior of the coefficients of the model. For example, the presence of common trends (or common cyclical components) in x_{nt} and $x_{n't}$ can be modelled by requiring that a peak in the spectral density appears at frequency zero (or at business cycle frequencies) and that the coherence at these frequencies is close to one. Following the same steps that led to (14)–(19), *a priori* beliefs concerning the presence of common stochastic trends can be summarized by means of the following three restrictions:

$$\sum_{j=0}^{N} \sum_{g=0}^{\infty} \eta_{njg} \frac{\sigma_{n,j}}{\sigma_{n,n}} \approx 0, \tag{26}$$

$$\sum_{j=0}^{N} \sum_{m=0}^{\infty} \eta_{n'jm} \frac{\sigma_{n'j}}{\sigma_{n'n'}} \approx 0, \tag{27}$$

$$\sum_{j=1}^{N} \left[\sum_{m=0}^{\infty} \eta_{n'jm} H_{2j} + \sum_{g=0}^{\infty} \eta_{njg} H_{1j} \right] \approx 0. \tag{28}$$

Hence, if in addition to the previous constraints one wants to build beliefs concerning the existence of common stochastic trends in the prior of the coefficients of x_{nt} and $x_{n't}$, one need only add $3M$ new rows to the matrix $\hat{R}$, and $3M$ new columns to r, where M is again the number of pairs to which the restriction applies. Once these restrictions are combined with the original DLS prior, they shrink the prior mean of the coefficients toward $\bar{r}$, rescale the elements of the main diagonal of the diagonal blocks (n, n, j) and (n', n', j), $j = 1, 2, \ldots, N$, and fill in the diagonal elements of the off-diagonal blocks (n, n', j) and (n', n, j) of the prior covariance matrix of the coefficients.

The restrictions implied by the presence of common patterns at cyclical frequencies are entirely analogous and, because of space limitations, will not be described here.

4. Inference

For given x_t and y_t, $\mathcal{L}(\theta|x_t y_t) \propto \int \mathcal{G}(x_t|y_t, \beta_t)\mathcal{F}(\beta_t|\theta)\,\mathrm{d}\beta_t$ is the marginalized likelihood function. If we summarize our prior views concerning the unknown θ through a density $\mathcal{Q}(\theta)$, then the posterior for θ is $\mathcal{P}(\theta|x, y) \propto \mathcal{L}(\theta|x, y)\mathcal{Q}(\theta)$. $\mathcal{P}(\theta|x, y)$ constitutes our source of inference concerning values and functions of θ. For the model under consideration very little is known about the properties of the θ's except, perhaps, the bounds of their support. Therefore, I assume that $\mathcal{Q}(\theta)$ is rectangular on a bounded subset of Θ. In this case $\mathcal{P}(\theta|x, y) \propto \mathcal{L}(\theta|x, y)$ and the marginalized likelihood function becomes the relevant source of inference of the θ's and for functions of them.

Doan, Litterman, and Sims (1984) showed that for each θ the marginalized likelihood can be computed as

$$\mathcal{L}(\theta|x_t, y_t) = \frac{T}{2} \times \log\left(\frac{1}{T}\sum_{t=1}^{T} \hat{u}_t(\xi_t)^{-1}\hat{\xi}\hat{u}_t'\right), \tag{29}$$

where $\hat{u}_t$ are the one-step-ahead recursive forecast errors and where $\hat{\xi}$ is the geometric mean (over t) of $\xi_t = \mathrm{var}_t(x_t)$. Note that a choice of θ which induces a $\mathcal{H}(x_t|y_t, \theta)$ with a large mass in a region where the data are unlikely to occur, will produce large forecast errors and therefore a low value of $\mathcal{L}$. Also, from (29) it is clear that only those forecast errors which occur when the covariance matrix of the coefficients is small significantly contribute to the likelihood. This scheme is particularly useful at the beginning of the estimation process when one-step-ahead forecast errors are large because coefficient uncertainty may be large. In this case the errors will receive less weight, leaving $\mathcal{L}$ relatively insensitive to alternative settings of the θ.

In evaluating the forecasts of the model, I will be interested in making probability statements regarding statistics of the forecasts of the model at various horizons. These statistics are functions $f(\theta)$ of the parameters of the model. Therefore the objects of interest are expressions of the form

$$\mathrm{E}(f(\theta)|\mathcal{A}, y, x) = \int_{\mathcal{A}} f(\theta)\mathcal{L}(\theta|x, y)\,\mathrm{d}\theta, \tag{30}$$

where $\mathcal{A} \subset \Theta$. Integrals like those appearing in (30) cannot be evaluated numerically using standard product or spherical rules when the dimension of

Θ is large. Naylor and Smith (1988), Niederreiter (1988), Geweke (1989), and Gelfand and Smith (1990), among others, describe ways of approximating them efficiently when simple quadrature rules are unfeasible. Here I will approximate (30) using the following procedure. First, I numerically evaluate the likelihood to find its maximum. To help to locate the peak of the likelihood function I employ the 'Bayesmth' algorithm written by Sims (1986) (details on the algorithm are provided in the appendix). Second, I choose $\mathscr{A}$ to be a hypercube of the parameter space which is centered around the vector of θ which maximizes the likelihood. The size and the exact shape of $\mathscr{A}$ will depend on the features of the likelihood function. Third, I randomly draw with replacement a large number of vectors $\bar{\theta}$ from $\mathscr{A}$. For each of them, I evaluate $\mathscr{L}(\bar{\theta}|x_t, y_t)$ using a Kalman filter algorithm and compute recursive forecasts $h(\hat{\beta}_t(\bar{\theta}))$ for steps 1 to 12 using recursively estimated $\hat{\beta}_t(\bar{\theta})$. Fourth, for each $\bar{\theta}$, I compute out-of-sample forecasting statistics $f(\bar{\theta}) = g(h(\hat{\beta}_t(\bar{\theta})))$. Fifth, I order the statistics and extract a 90% confidence band around the mean value of the statistics.

5. An example

In this example I employ quarterly data on total industrial production (IP) for three European countries: France, West Germany, and Italy. I chose to keep N small for computational purposes and hardware limitations. These countries were selected for two reasons. First, because their total industrial production indices measure comparable aggregates over time.[3] Second, because of their geographical proximity and close economic links, these aggregates are likely to form a group with strong common features.

5.1. The data

The data is taken from the OECD Main Economic Indicators for the period 1960.1–1978.4 and from the OECD Indicators of Industrial Activities for the period 1979.1–1989.2 and converted to 1980 = 100 indices. The resulting quarterly time series are plotted in fig. 1. The estimated log spectral density for each series after deterministic seasonals are extracted using seasonal dummies appears in the left panel of fig. 2. The estimated coherence for each pair of series are presented in the right panel of fig. 2.

Inspection of fig. 1 reveals some interesting features of the data. First, a noticeable break in the growth pattern of all series appears in 1975. Second, seasonals are not very regular. For example, the French IP series has a mild

[3] Up to 1977.4 none of these indices include construction. Also, contrary to indices for the U.S., Japan, and Canada, the quarterly IP indices used here measure at least 75% of the value added by total industrial activities in the country.

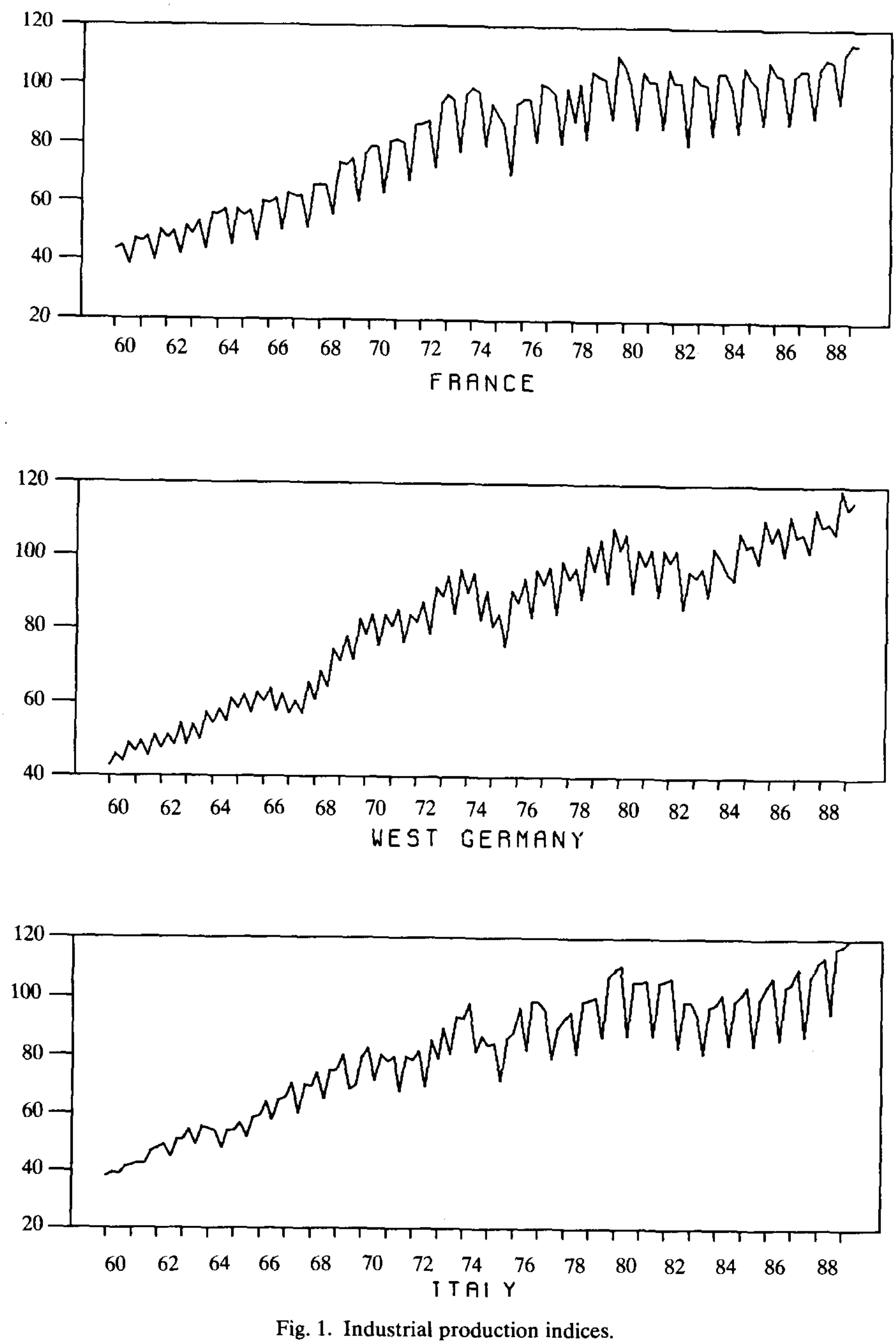

Fig. 1. Industrial production indices.

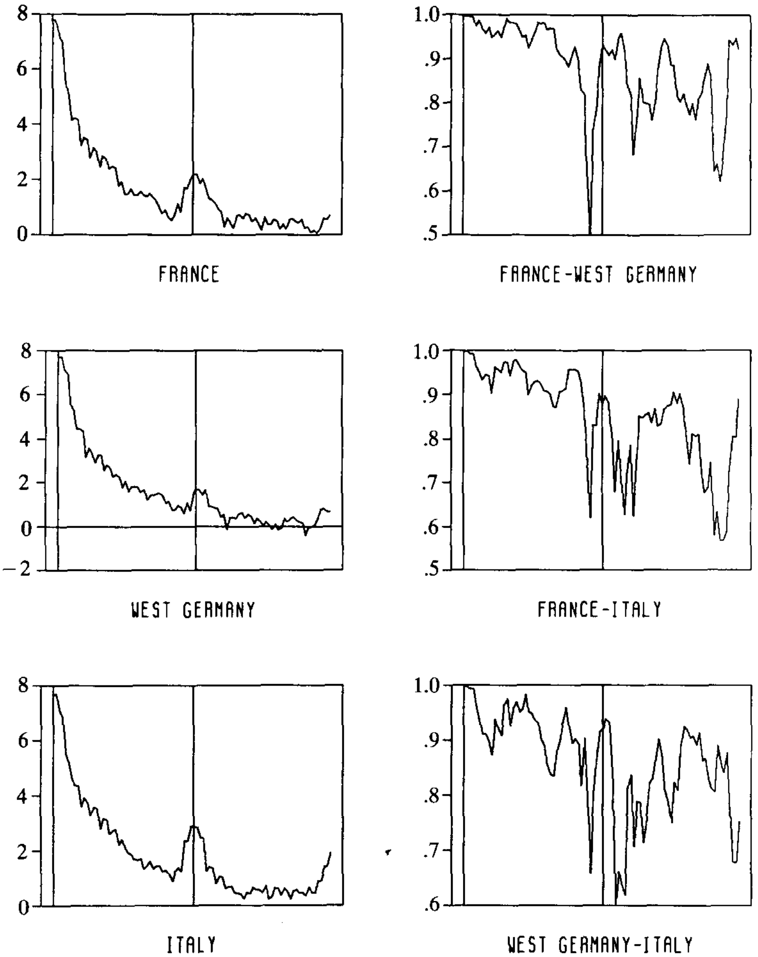

Fig. 2. Estimated log spectra and coherences.

six-month pattern in the first five years of the data, which disappears up to 1978 and reappears again with modifications in the 80's. The Italian IP series is the most irregular one. Seasonals seem to change structure every four or five years, and during the mid 70's the pattern is buried in the cyclical downturn of the series. These observations suggest that seasonals will show up as a large mass, as opposed to a sharp peak, in the spectral density of the series. This is confirmed in fig. 2, which also shows that all IP series display significant variations at $\pi/2$ (one-year cycle) which are not entirely deterministic. In addition, fig. 2 indicates that common patterns do exist at frequencies 0 and $\pi/2$ for all pairs but not at π (two-quarter cycles).

I use data from 1960.1–1981.4 to construct the likelihood function of the θ's.[4] Eight lags of each variable are used in the model. The 'start-up' values for the coefficients of the model are given by $\beta_1 \sim \mathcal{N}(\beta_0, (1-\theta_0)^{-2} \times \operatorname{var}(\beta_0))$. Forecasts are computed recursively from 1982.1 to 1989.2 using updated β_t vector, for a given $\bar{\theta}$ vector. The 90% range around the mean of each of the forecasting statistics is computed using the procedure described in the previous section.

5.2. *The forecasting exercise*

I am primarily interested in evaluating the contribution of various parameters to the forecasting performance of the model. This exercise is useful since the paper adds several features to the standard DLS procedure and it is of interest to evaluate which of these new features, if any, produce improvements in the forecasting performance of the model.

I conduct the exercise in two ways. First, since the prior of the model is formulated very generally and various specifications are nested in it, it is possible to trace out the contribution of each additional feature to the forecasts of the model by simply conditioning the construction of the posterior distribution of the parameters over certain dimensions and compute the statistics of interest using this conditional posterior distribution. Let $\theta = [\gamma_1, \gamma_2]$ be a partition of the parameters of the model and let $L = \mathcal{L}(\gamma_1 | x_t, y_t, \gamma_2 = \bar{\gamma}_2)$ be the likelihood function of γ_1 conditional on γ_2 taking a specific value. Then in each of these exercises I construct a 90% confidence range around the mean of each Theil U-statistic using (30) with L in place of $\mathcal{L}$ and choose $\mathcal{A}$ to be the hypercube around each $\max_{\gamma_1} \mathcal{L}(\gamma_1 | x_t, y_t, \gamma_2 = \bar{\gamma}_2)$. I also compare the peak of the likelihood for various partitions of θ into γ_1 and γ_2 using the Schwarz (1978) criterion.

Second, since there is evidence that *any* shrinkage procedure improves the forecasts of a model which has a large number of parameters [see, e.g., Thisted and Wecker (1981), Garcia-Ferrer, Highfield, Palm, and Zellner (1987)], it is of interest to know whether or not a model constrained with the common pattern restrictions improves upon the forecasts of a model which is constrained by arbitrary restrictions.[5] To do this I arbitrarily chose the parameters of the DLS-type prior and make no attempt to fine tune them.

[4] I chose 1981.4 as the ending date of the estimation because the features of all four series are altered after that date, and it is of interest to study how the model performs when faced with a possible structural change.

[5] I would like to thank J. Geweke for suggesting this type of exercise.

Then I compare the 90% confidence range around the mean of each Theil U-statistic of this rough DLS prior shrunk with the common pattern restrictions to the 90% confidence range around the mean of each Theil U-statistic of the same rough DLS prior shrunk with arbitrary restrictions.

For the first experiment, I examine five different specifications. The most restrictive one is a model where seasonality is accounted for by seasonal dummies, no frequency domain restrictions are used, and the prior on the coefficients is made uninformative. This is achieved by setting $\theta_7 = \theta_8 = 0$, $\theta_2 = \theta_9 = 10^6$, and $\theta_4 = \theta_5 = 1$, $\forall n, n', k$. This setup approximates an unrestricted panel-VAR model with constant coefficients and seasonal dummies (model I in the tables). The second most restricted specification still accounts for seasonals with dummies, imposes no frequency domain restrictions, but allows a prior on the coefficients similar to a standard DLS form. Here I restrict $\theta_4 = 1$, $\theta_8 = 0$, $\theta_9 = 10^6$ (model II), but contrary to the standard DLS prior I allow $\theta_7 \neq 1$. Since the prior on the coefficients of the dummies in each equation is the same, this setup mimics a situation where common seasonals are accounted for by an exchangeable prior on the seasonal dummies. The third model is similar to the above second specification, but allows the prior mean and the prior variance of the AR coefficients to have a seasonal structure. In this case I only restrict $\theta_9 = 10^6$ (model III). The fourth specification is similar to the third one, but adds the common stochastic seasonal restriction at both seasonal frequencies (model IV in the tables). The final specification adds to the fourth one the common stochastic trend restrictions (model V).

For the second experiment, I first compute baseline point forecasts with a roughly chosen DLS prior (model VI). Then, adding as a shrinking element the batting average of the Minnesota Twins for the 1989–89 season, I recompute the forecasts for the model (model VII) and compare them with the forecasts obtained by introducing the common stochastic seasonal restrictions into the rough DLS prior (model VIII).

Table 1 reports the parameter vector that produces the peak of the likelihood function for the various specifications and the upper and lower limits of the hypercube $\mathscr{A}$ for each dimension. Table 2 reports the 90% range around the mean Theil U-statistic at one, four, and eight quarters ahead, the 90% range for five average Theil U-statistics (one over twelve steps, three over variables at one, four, and eight steps, and one overall average over steps and variables) for each of the eight specifications and the value of the Schwarz information criteria.[6] For comparison columns 9 to 11 of table 2 also report point estimates of the same forecasting statistics

[6]On a PC 386, 20MHZ machine with 387 math coprocessor and no extended memory, the total computation time for table 2 was about 480 hours.

Table 1
Optimal value of the parameters and range of $\mathscr{A}$.

Parameter	Models							
	VIII	VII	VI	V	IV	III	II	I
θ_0	1.00	1.00	1.00	0.999 [0.98, 1.00]	0.998 [0.98, 1.00]	0.998 [0.98, 1.00]	0.99 [0.96, 1.00]	0.995 [0.97, 1.00]
θ_1	0.00	0.00	0.00	3.06 [2.08, 4.07]	3.21 [2.33, 4.02]	3.35 [2.44, 4.18]	2.86 [2.04, 3.85]	0.00
θ_2	0.20	0.20	0.20	0.0002 [0.01, 0.0001]	0.0004 [0.01, 0.0001]	0.0003 [0.01, 0.0001]	0.50 [0.20, 0.80]	1.00
θ_3	2.00	2.00	2.00	0.01 [0.0001, 0.05]	0.01 [0.0001, 0.05]	0.01 [0.0001, 0.05]	0.005 [0.0001, 0.02]	0.01 [0.0001, 0.05]
θ_4	1.00	1.00	1.00	197 [151, 259]	198 [151, 259]	204 [173, 265]	1.0	1.0
θ_5	0.50	0.50	0.50	0.21 [0.11, 0.28]	0.18 [0.09, 0.25]	0.14 [0.07, 0.22]	0.28 [0.15, 0.33]	0.35 [0.27, 0.42]
θ_6	6.00	6.00	6.00	181 [80, 250]	174 [60, 200]	197 [90, 250]	214 [100, 250]	1.0

θ_7	1.00	1.00	1.00	0.001 [0.0001, 0.06]	0.001 [0.0001, 0.06]	0.001 [0.0001, 0.06]	0.95 [0.77, 0.99]	1.0
θ_8	0.00	0.00	0.00	1.44 [1.20, 1.60]	1.43 [1.20, 1.60]	1.35 [1.15, 1.55]	0.0	0.0
θ_9	5.20 [3, 10]		10^6	1.1 [0.4, 1.5]	3.4 [2.0, 8.0]	10^6	10^6	10^6
θ_{Twins}		0.102 [0.90, 1.05]						
$\sigma_{z_{11}}$				2.9 [1.8, 3.8]	10^5			
$\sigma_{z_{22}}$	1.0	1.0	1.0	1.0	1.0			
$\sigma_{z_{33}}$				10.78 [7.0, 19.0]	5.86 [3.0, 8.0]			
$\sigma_{z_{44}}$				0.76 [0.40, 2.98]	10^5			
$\sigma_{z_{55}}$				0.34 [0.13, 0.73]	1.02 [0.81, 1.25]			
$\sigma_{z_{66}}$				4.47 [2.50, 6.50]	10.39 [8.0, 15.0]			

Table 2

Theil U-statistics; 90% range around the mean value; forecasting sample 1982.1–1989.12.[a]

		Models																	
	Step	VIII		VII		VI	V		IV		III		II		I		ARIMA	AR(8)	TS
France	1	0.30	0.33	0.32	0.36	0.33	0.14	0.18	0.15	0.18	0.16	0.19	0.21	0.27	0.28	0.35	0.19	0.18	0.83
	4	2.40	2.46	2.66	2.81	2.45	0.89	0.99	0.93	1.00	0.95	1.02	1.19	1.68	1.25	1.66	0.99	1.14	6.21
	8	1.94	2.08	2.07	2.35	2.07	0.83	0.92	0.86	0.96	0.90	1.02	1.00	1.10	0.88	1.19	1.13	1.04	5.14
	Avg. 12	0.80	0.83	0.85	0.90	0.89	0.34	0.37	0.36	0.39	0.38	0.42	0.49	0.59	0.58	0.73	0.49	0.43	2.03
WG	1	0.62	0.65	0.68	0.71	0.69	0.48	0.53	0.51	0.54	0.56	0.59	0.65	0.74	0.47	0.52	0.48	0.46	1.24
	4	1.17	1.25	1.45	1.57	1.21	0.84	0.90	0.83	0.94	0.88	0.99	0.95	1.08	1.36	1.59	0.81	1.13	2.22
	8	1.17	1.24	1.16	1.26	1.45	0.67	0.73	0.69	0.74	0.77	0.80	0.77	0.82	1.38	1.63	0.68	0.76	1.49
	Avg. 12	0.84	0.88	0.91	0.97	0.89	0.61	0.67	0.63	0.69	0.68	0.77	0.73	0.82	0.96	1.12	0.57	0.68	1.29
Italy	1	0.33	0.36	0.32	0.37	0.33	0.27	0.38	0.29	0.38	0.30	0.40	0.35	0.46	0.27	0.40	0.35	0.30	0.99
	4	0.92	0.95	0.95	1.01	0.95	0.90	0.95	0.92	1.00	0.93	1.00	1.12	1.26	1.42	1.66	1.15	1.33	2.66
	8	1.12	1.23	1.15	1.26	1.23	0.66	0.76	0.68	0.77	0.71	0.75	0.76	0.80	0.90	1.03	1.17	1.18	4.06
	Avg. 12	0.49	0.53	0.50	0.52	0.52	0.43	0.48	0.46	0.49	0.47	0.49	0.50	0.56	0.69	0.90	0.72	0.76	1.61
	Avg. 1	0.416	0.446	0.440	0.480	0.450	0.296	0.363	0.316	0.366	0.340	0.393	0.403	0.490	0.346	0.423	0.340	0.313	1.020
	Avg. 4	1.496	1.553	1.686	1.796	1.536	0.876	0.946	0.893	0.980	0.920	1.003	1.086	1.340	1.343	1.636	0.952	1.200	3.030
	Avg. 8	1.410	1.516	1.460	1.623	1.583	0.720	0.803	0.743	0.823	0.793	0.856	0.843	0.906	1.053	1.263	0.950	0.993	3.563
	Ov. ave.	0.710	0.746	0.753	0.796	0.766	0.460	0.506	0.483	0.523	0.510	0.560	0.573	0.656	0.743	0.947	0.593	0.623	1.643
Schwarz		1245		1260		1263	2004.3		1987.5		1956.7		1892.0		1450.1				

[a]Model I approximates an unrestricted VAR model with seasonal dummies; model II is a model with DLS prior and an exchangable prior on the seasonal dummies; model III adds to model II a seasonal prior mean and seasonal prior variance; model IV adds to model III the common seasonal restriction; model V adds to model IV the common trend restriction; model VI has no seasonal dummies and a rough and ready DLS prior; model VII has no seasonal dummies, a rough and ready DLS prior, and an arbitrary shrinkage prior (Minnesota Twins batting average); model VIII has no seasonal dummies, a rough and ready DLS prior, and a seasonality-based shrinkage prior. The ARIMA model is the same for each series and has the format $(1-L^4)y_t=(1-\alpha L)\varepsilon_t$. AR(8) is an unrestricted AR model with eight lags and no prior. TS indicates a model with a linear time trend and seasonal dummies.

obtained from a multiplicative ARIMA model,[7] from a univariate unrestricted AR(8) model, and from a univariate model with a linear time trend and deterministic seasonal dummies.

5.3. The results

Before examining the results of the forecasting exercise I will briefly describe some features of the likelihood function for this problem. First, although the likelihood is relatively insensitive to alternative choices of parameters in several dimensions, it has very narrow contours in the θ_0, θ_1, θ_4, and θ_8 dimensions. The variations in the forecasting statistics reported in table 2 are primarily due to variations in these four parameters. Second, the likelihood function has a very steep ridge in the region of nonstationary values for the prior mean of the fourth lag coefficient of each equation of the model. For example, if we restrict θ_8 to lie in the $[0, 1]$ interval, the value of the log likelihood drops by about 400 points. Third, conditioning on $\theta_3 = \theta_7 = 0$ does not affect the maximized value of the likelihood. Therefore, these two parameters can be treated as fixed and eliminated from the maximization process. Fourth, the value of θ_5 which maximizes the likelihood is low, and there is a rapid deterioration in the forecasting performance for values of θ_5 close to zero or higher than 0.2. Finally, as expected from the discussion of figs. 1 and 2, the likelihood function is relatively flat for large values of the variance of the common seasonal restrictions and drops sharply as these parameters are driven toward zero.

Several important features emerge from table 2. Considering first the overall average Theil U-measure (last row of the table), one can see that a finely tuned DLS prior with an exchangable prior on the seasonal dummies is inadequate for these three seasonal series. For example, the overall performance of this model is unimpressive when compared to that of ARIMA models. A prior mean different from zero on the fourth lag and a prior covariance matrix which allows seasonal lags to be substantially larger than the others seem necessary to forecast these IP series. Accounting for the presence of common stochastic patterns at seasonal frequencies improves the forecasting performance of the model as long as the restriction is not too tightly imposed. Similarly, taking into account the presence of common patterns at the zero frequency produces gains, especially in the long run. The gains in the short run, however, are not too significant. Fig. 3 presents the actual data and a 90% confidence interval for the forecasts generated by model V for the period 1986.3–1989.2 using coefficient estimates obtained

[7]The ARIMA model is the same for each series and is of the form $(1 - L^4)y_t = (1 - \alpha L)\varepsilon_t$. Among alternative specifications producing very similar autocovariance function, this was chosen because it produced the best forecasting performance as measured by the one-step-ahead Theil U-statistic for the period 1982–89.

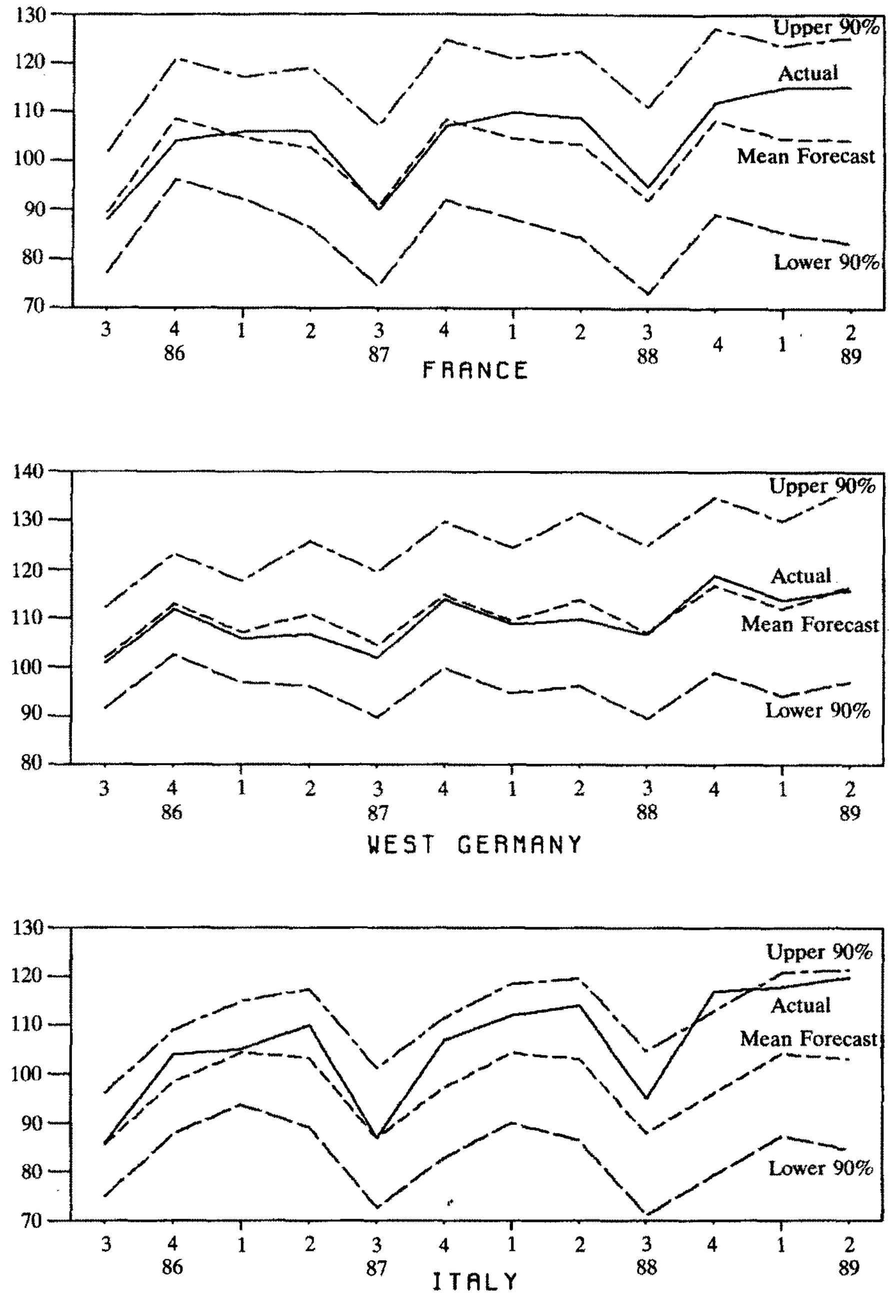

Fig. 3. Forecasts of model V.

with information up to 1986.2. It is clear from the figure that the mean forecasts of this model track the actual data pretty well.

When we look at individual series and horizons, the results are more mixed. The common seasonal restrictions appear to be very useful in forecasting the Italian IP index at all horizons, but not the West German IP index. The loss in forecasting performance for this series is significant particularly at the one-step-ahead horizon. On average, the restrictions appear to be important in improving the forecasting performance at seasonal and longer horizons for all series.

Table 2 also indicates that the improvements obtained by introducing the common seasonal restriction are larger on average than those obtained by simply shrinking a rough 'Litterman' prior with an arbitrary restriction. The gains are not astonishingly large however and, on average, they are of the order of 4% to 5% points. In addition, since at shorter horizons the model shrunk with an arbitrary restriction is as good as the model shrunk with the common pattern restriction, no strong conclusions regarding the usefulness of the common seasonal restrictions can be drawn from the experiment.

Table 2 finally shows that the model under consideration improves upon the forecasts of other standard specifications. The gains are particularly evident in the long run [see, in particular, the average (over variables) Theil U at eight steps ahead]. Using the overall average (over steps and variables) Theil U as a term of comparison, I find that the improvement over the ARIMA model and the AR(8) model is of the order of 20% to 25% points. At one step ahead, on the other hand, the performance is series dependent. On average at one step ahead the forecasts of the model are not too different from those of the ARIMA model and the univariate AR(8) model.[8]

6. Conclusions

This paper presents an alternative methodology for modelling and forecasting series which possess common patterns at seasonal frequencies. The approach is in the Bayesian autoregressive tradition pioneered by Doan, Litterman, and Sims (1984) and Sims (1989) and builds the presence of common deterministic and stochastic patterns directly into the prior of the coefficients of the model. While common deterministic patterns are accounted for with an exchangable prior on the coefficients of the deterministic

[8] In a previous version of the paper I also examined the sensitivity of the forecasting statistics to changes in the sample used to construct the posterior of the parameters. The hope is that the features of the series represented by the θ's are time-invariant so that these parameters need not be reestimated very often. For this exercise I used three subsamples (1960–72, 1960–77, 1960–81) and computed forecasting statistics over the 1982–1989 period using the optimal parameters obtained in each subsample. The results indicated that, although the optimal parameters of the model do not change very much over the three subsamples chosen, the forecasts of the three specifications differed substantially.

variables, *a priori* beliefs concerning the existence of common stochastic features at certain frequencies are captured with a set of uncertain linear restrictions on the AR coefficients of a vector of time series. These restrictions are combined with a standard Doan–Litterman–Sims prior to produce a revised prior for the coefficients of the model which has the common pattern restrictions built in it.

As an illustration the proposed methodology is used to model common patterns at seasonal and zero frequencies for a small panel of industrial production indices, and the forecasting performance of various model specifications is compared with standard univariate specifications. The results indicate that the technique is flexible in adapting to situations where the underlying economic structure frequently changes and that accounting for common patterns in the prior of the coefficients is potentially useful in improving the forecasts of the model.

Appendix

This appendix briefly describes the features of the 'Bayesmth' algorithm which is employed to locate the peak of the likelihood function. For more details the reader should consult Sims (1986).

The algorithm takes as input a set of values for the parameter vector θ and for the likelihood function $\mathscr{L}$ $(\theta_i, \mathscr{L}_i,\ i = 1, 2, \ldots, n)$ and constructs an interpolated function $\mathscr{L}^n$, satisfying $\mathscr{L}^n$ $(\theta_i = \mathscr{L}(\theta_i),\ \forall i)$ using either a third-order cubic spline or a Gaussian kernel. Once the interpolated function has been constructed, the algorithm numerically searches for the maximum of $\mathscr{L}^n$.

In the batch version of the program used here (the LOOPSMTH routine), the algorithm returns the value of θ corresponding to the guessed maximum value of $\mathscr{L}^n$. Then, treating this value for the vector θ as the $n+1$ observation and the corresponding value of $\mathscr{L}(\theta_{n+1})$ as $\mathscr{L}_{n+1}$, the algorithm repeats the interpolation to obtain a function $\mathscr{L}^{n+1}$, searches for the maximum of $\mathscr{L}^{n+1}$, and returns the value of θ corresponding to the guessed maximum value of $\mathscr{L}^{n+1}$.

The searching procedure for the maximum stops when $|\max_\theta \mathscr{L}^{n+k}(\theta) - \max_\theta \mathscr{L}^{n+k-1}(\theta)| < \varepsilon$, where ε is a prespecified and appropriately chosen constant.

References

Burns, A. and W. Mitchell, 1946, Measuring business cycles (National Bureau of Economic Research, New York, NY).

Canova, F., 1992, An alternative approach to modelling and forecasting seasonal time series, Journal of Business and Economic Statistics 10, 97–108.

Carlin, J.B., A.P. Dempster, and A.B. Jonas, 1985, On the model and methods for Bayesian time series analysis, Journal of Econometrics 30, 67–90.
Doan, T., R. Litterman, and C. Sims, 1984, Forecasting and conditional projections using realistic prior distributions, Econometric Reviews 3, 1–100.
Drèze, J., 1975, Bayesian theory of identification in simultaneous equation models, in: S. Fienberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics (North-Holland, Amsterdam).
Engle, R. and C. Granger, 1987, Cointegration and error correction: Representation, estimation and testing, Econometrica 55, 251–276.
Garcia-Ferrer, A., R.A. Highfield, F. Palm, and A. Zellner, 1987, Macroeconomic forecasting using pooled international data, Journal of Business and Economic Statistics 5, 53–67.
Gelfand, A. and A. Smith, 1990, Sampling based approaches to calculating marginal densities, Journal of the American Statistical Association 85, 398–409.
Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, Econometrica 57, 1317–1340.
Ghysels, E., 1988, A study toward a dynamic theory of seasonality for economic time series, Journal of the American Statistical Association 82, 168–172.
Gourieroux, C. and I. Paucelle, 1989, Detecting a long run relationship (with application to PPP hypothesis), Working paper no. 8902 (CEPREMAP, Paris).
Granger, C., 1979, Seasonality: Causation, interpretation and implications, in: A. Zellner, ed., Seasonal analysis and economic time series (U.S. Department of Commerce, Bureau of Census, Washington, DC).
Granger, C. and A. Weiss, 1983, Time series analysis of error correction models, in: Karlin, Samad, T. Amemiya and P. Goodman, eds., Studies in econometrics (Academic Press, New York, NY).
Hansen, L.P. and T. Sargent, 1990, Recursive linear models of dynamic economies (Princeton University Press, Princeton, NJ).
Hylleberg, S. 1986, Seasonality in regression (Academic Press, New York, NY).
Hylleberg, S., R. Engle, C. Granger, and S. Yoo, 1990, Seasonal integration and cointegration, Journal of Econometrics 44, 215–238.
Kang, H., 1988, Common deterministic trends, common factors and cointegration, Manuscript (University of Indiana, Bloomington, IN).
Lindley, A. and A. Smith, 1972, Bayes estimates for the linear model, Journal of the Royal Statistical Society B 34, 1–41.
Litterman, R., 1980, Techniques for forecasting using vector autoregressions, Unpublished Ph.D. thesis (University of Minnesota, Minneapolis, MN).
Litterman, R., 1986, Forecasting with Bayesian VAR: Four years of experience, Journal of Business and Economic Statistics 4, 25–38.
McNees, S., 1986, Forecasting accuracy of alternative techniques: A comparison of US macroeconomic forecasts, Journal of Business and Economic Statistics 4, 5–23.
Monahan, J., 1983, Fully Bayesian analysis of ARMA time series models, Journal of Econometrics 21, 307–321.
Naylor, J. and A. Smith, 1988, Econometric illustrations of novel numerical integration strategies for Bayesian inference, Journal of Econometrics 38, 103–125.
Niederreiter, H., 1988, Quasi Monte Carlo methods for multidimensional numerical integration, International Series of Numerical Mathematics 85, 157–171.
Quah, D., 1989, International patterns of growth: I. Persistence in cross-country disparities, Manuscript (Massachusets Institute of Technology, Cambridge, MA).
Raynauld, J. and J.-G. Simonato, 1993, Seasonal BVAR models: A search along some time domain priors, Journal of Econometrics, this issue.
Rozanov, Y.A., 1967, Stationary random processes (Holden-Day, San Francisco, CA).
Schwarz, G., 1978, Estimating the dimension of a model, American Statistician 6, 461–464.
Sims, C., 1979, Comment to analysis and modelling of seasonal time series, in: A. Zellner, ed., Seasonal analysis of time series (U.S. Department of Commerce, Bureau of Census, Washington, DC) 472.
Sims, C., 1986, Bayesmth: A computer program for Bayesian interpolation, Working paper no. 234 (University of Minnesota, Minneapolis, MN).

Sims, C., 1989, Nine variable probabilistic macroeconomic forecasting model, Discussion paper no. 14 (Federal Reserve of Minneapolis, Institute for Empirical Macroeconomics, Minneapolis, MN).

Stock, J. and M. Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.

Thisted, R. and W. Wecker, 1981, Predicting a multitude of time series, Journal of the American Statistical Association 76, 516–523.

Zellner, A., 1971, Introduction to Bayesian inference in econometrics (Wiley, New York, NY).

Zellner, A. and H.S. Hong, 1989, Forecasting international growth rates using Bayesian shrinkage and other procedures, Journal of Econometrics 40, 183–202.