Econometric Theory, **10**, 1994, 645-671. Printed in the United States of America.

WHAT MACROECONOMISTS SHOULD KNOW ABOUT UNIT ROOTS

A Bayesian Perspective

HARALD UHLIG
Center for Economic Research
Tilburg University

This paper summarizes recent Bayesian research on unit roots for the applied macroeconomist in the way Campbell and Perron [8] summarized the classical unit roots perspective. The appropriate choice of a prior is discussed. In recognizing a consensus distaste for explosive roots, I find the popular Normal-Wishart priors centered at the unit root to be reasonable provided they are modified by concentrating the prior mass for the time trend coefficient toward zero as the largest root approaches unit from below. I discuss that the tails of the predictive density can be sensitive to the prior treatment of explosive roots. Because the focus of an investigation often is on a particular persistence property or medium-term forecasting property of the data, I conclude that Bayesian methods often deliver natural answers to macroeconomic questions.

1. INTRODUCTION

Campbell and Perron [8] recently summarized much of the research on classical unit root econometrics. Their rules or observations are partly guidelines on how to proceed and partly warning signs on what to avoid. This paper formulates additional observations for the applied macroeconomist from a Bayesian perspective for two reasons. First, classical and Bayesian inference on unit roots can differ substantially. Second, a Bayesian perspective delivers natural answers by allowing the researcher to state the uncertainty in answering a particular question without first having to take a stand on whether the data is integrated or trend stationary.

I am indebted to Chris Sims, who taught me much of what went into this paper but may disagree with its contents. I am grateful to many participants at the Bayes Methods and Unit Roots Conference at Yale, April 1992, and for participants at a seminar in macroeconomics at Princeton University for helpful comments and discussions, in particular John Campbell, Gregory Chow, John Cochrane, Herman van Dijk, Pierre Perron, Peter Phillips, Albrecht Ritschl, Peter Schotman, and James Stock. I am grateful to the hospitality of the SFB 303 at the University of Bonn, where I completed this research.

 0266-4666/94 $5.00 + .00

Macroeconomists are often interested in some particular persistence property or medium-term forecasting property of the data. Roots near unity are important only insofar as they sometimes imply sizable persistence of shocks,[1] as has been argued by, e.g., Christiano and Eichenbaum [11], Cochrane [12–14], Durlauf [24], and Sims [96]. To a Bayesian, a root of unity or a particular trend-stationary specification is one of several possibilities and receives posterior weight according to the evidence in the data. By using these weights, the macroeconomic question of interest can be answered without the need for pretesting.

After a review of Bayesian methodology for autoregressive time series in Section 2, I discuss the choice of priors for the univariate case in Section 3. That section is more technical than the rest of the paper. The fragility of some inferences to the prior treatment of explosive roots is discussed. Recognizing a consensus distaste for explosive roots, the popular Normal–Wishart priors centered at the unit root are reasonable provided they are modified by concentrating the prior mass for the time trend coefficient toward zero as the largest root approaches unity from below. In turning to the persistence issue in Section 4, I discuss how Bayesian methods naturally take account of the parameter uncertainty, and that the tails of the predictive density can be sensitive to the prior treatment of explosive roots.

2. BAYESIAN METHODS

For a thorough introduction, see Box and Tiao [2], Zellner [106], Leamer [51], Judge et al. [37], Berger [3], Broemeling [5], West and Harrison [105], and Lütkepohl [62]. For a discussion of methodological issues, see Hendry and Richard [32], Hendry [33], Leamer [54], Pagan [66,67], Phillips [69], Sims [92,94], Berger and Wolpert [4], Poirier [77,78], Rust [81], Geweke [28], Zellner [108], Pagan and Wickens [68], and Kydland and Prescott [50]. Consider the m-variate model,

$$Y_t = B_{(1)}Y_{t-1} + B_{(2)}Y_{t-2} + \cdots + B_{(k)}Y_{t-k} + C_{(0)} + C_{(1)}t + \epsilon_t, \tag{1}$$

where $t = 1, \ldots, T$ denotes time, where the data vectors Y_t, $t = 1-k, \ldots, T$ are of size $m \times 1$, where the coefficient matrices $B_{(i)}$, $i = 1, \ldots, k$ are of size $m \times m$, where $C_{(0)}$ and $C_{(1)}$ are of size $m \times 1$, and where ϵ_t, $t = 1, \ldots, T$, size $m \times 1$, are independently and normally distributed[2] according to

$$\epsilon_t \sim \mathcal{N}(0,\Sigma)\text{i.i.d.}, \; t = 1, \ldots, T. \tag{2}$$

What is known at inference time is the data. What is unknown are the regression coefficients, summarized in $B = [B_{(1)}\, B_{(2)} \cdots B_{(k)}\, C_{(0)}\, C_{(1)}]'$, and the covariance matrix Σ or, alternatively, the precision matrix $H = \Sigma^{-1}$. Whereas a classical approach conditions on some particular values for the parameters and calculates the sampling distribution of the data, Bayesian inference conditions on the observed data and calculates probability distri-

butions for the parameters. A Bayesian first chooses a prior probability density function $\pi(B,H)$ in B and H: more on that below. By multiplying this prior with the likelihood function, one obtains the posterior

$$\pi_T(B,H) \propto \pi(B,H)L(B,H; Y_{-k+1}, \ldots, Y_t) \tag{3}$$

with the constant of proportionality so that $\pi_{T,H}$ integrates to unity. The posterior is the key tool and is used to answer the particular question at hand; see Section 4. These questions often take the form of evaluating the posterior expectation $E_T[g(B,H)]$ for some function $g(\cdot,\cdot)$ of interest, where the expectation is taken with respect to the posterior. It is therefore useful to report this expectation along with its variance $\mathrm{Var}_T[g(B,H)]$ and show the sensitivity of these answers to choosing among several reasonable priors.

One may think about calculating $\pi_T(B,H)$ in two steps: first, find the "prior" $\pi_0(B,H)$ given the initial observations,

$$\pi_0(B,H) \propto \pi(B,H)L(B,H; Y_{-k+1}, \ldots, Y_0), \tag{4}$$

and then calculate the posterior $\pi_T(B,H)$ by using the likelihood function conditional on the initial observations,

$$\pi_T(B,H) \propto \pi_0(B,H)L(B,H; Y_1, \ldots, Y_t | Y_{-k+1}, \ldots, Y_0). \tag{5}$$

For the model in (1) and (2), the conditional likelihood function that appears in (5) turns out to have a particularly simple form (for a proof, see Appendix).

OBSERVATION 1. *Given the data Y_t, $t = -k+1, \ldots, T$, the conditional likelihood function as a function in B and H is proportional to a Normal-Wishart*[3] *density function. This is true regardless of whether there are unit roots, cointegrating vectors, explosive roots, or not.*

This observation is the central message in Sims and Uhlig [97]: whereas the conditional likelihood function viewed as a function of the data given the parameters may not be standard (the classical perspective), the conditional likelihood function viewed as a function of the parameters given the data is standard (the Bayesian perspective). Crudely speaking, the nonstationarity is in the data, not in the parameters, and the data is given at inference time. By conditioning on the covariance matrix Σ as well, the shape of the likelihood function is simply proportional to a normal distribution and can therefore be summarized by its mean and variance.

OBSERVATION 2. *Conventional t and F statistics and their conventional p-values are meaningful in summarizing the shape of the likelihood function, regardless of whether there are unit roots or not.*

Most macroeconomists may feel uncomfortable applying Bayesian methods because there seems to be too much choice in specifying a prior. In practice, however, only a few candidates are actually useful and used. There are

up to four reasons to choose a particular prior: the prior is generally agreed upon for reporting results, the prior is convenient for calculating results, the prior expresses "ignorance," and the prior expresses subjective prior beliefs about the parameters of interest.

Remarkably enough, in linear models similar to (A.1) with exogeneous, rather than endogeneous, regressors, these four reasons are rarely in conflict, and it has become standard practice to choose a "flat prior," i.e., a prior proportional to $|H|^{-(m+1)/2}$ (see, e.g., Zellner [106, Sect. 8.1]), or, more generally, a Normal-Wishart prior[4] in B and H. The posterior will then be Normal-Wishart as well. This result is still true in the models in (1) and (2) because Bayesian inference is conditional on the observed data (see Appendix for the precise formulas).

OBSERVATION 3. *If the prior π_0 is given by a Normal-Wishart density, then the posterior π_T is given by a Normal-Wishart density as well.*

This observation explains the popularity of the Normal-Wishart prior. Unfortunately, and in contrast to the situation with exogeneous regressors, flat priors or, more generally, Normal-Wishart priors, can be quite informative about certain properties of the autoregressive model above. This was pointed out by Phillips [70] and discussed by Koop and Steel [43], Leamer [55], Kim and Maddala [38], Poirier [79], Schotman and van Dijk [85], Stock [99], DeJong and Whiteman [17], Sims [98], Phillips [71], and Schotman [87]. Briefly, the argument is that parameter regions, where the data will easily distinguish nearby values, are packed "denser" than others. To find an uninformative prior, it may therefore be sensible to reparameterize the model so that the parameter space becomes "evenly packed" as measured by the data to be observed. In other words, one may desire to reparameterize the model so that the Fisher information matrix function I is constant on the new parameter space and impose a flat prior there. By using calculus, this amounts to a prior proportional to $|I|^{1/2}$ in the original parameterization. Such a prior assigns a high density to parameter regions where the data should be very informative. This "uninformative prior" is called the Jeffreys prior, after Jeffreys [35,36]. Jeffreys prior has the additional and sometimes desirable property that it is immune to reparameterizations. For simple, univariate models, Jeffreys prior will be calculated and discussed in Section 3 below. Although few endorse the uncritical use of Jeffreys prior, it provides at least a helpful benchmark.

OBSERVATION 4. *Calculating Jeffreys prior is helpful in choosing a suitable prior and understanding its implications.*

Another candidate for priors representing "knowing little" are the maximal data information prior (MDIP) distributions; see Zellner [106,107,110] and Zellner and Min [112]. They are given by $\pi(\theta) \propto \exp(\iota(\theta))$, where $\iota(\theta) = \int p(y|\theta)\log(p(y|\theta))\,d\theta$ is Shannon's [89] measure of the informa-

tion in the data distribution and where $p(y|\theta)$ is the probability density for the data y given the parameter vector θ. The MDIP is not invariant to reparameterizations.

3. PRIORS FOR UNIVARIATE MODELS

The recent debate has centered on the difference between Jeffreys prior and priors like the Normal–Wishart prior in the univariate case (see Phillips [70] and his discussants)[5] and is summarized below. The focus here is on the reasonableness and potential modifications of the popular Normal–Wishart prior. This section is more technical than the rest of the paper.

3.1. Excluding a Trend

Consider the simple AR(1) model

$$y_t = \rho y_{t-1} + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0,\sigma^2)\text{i.i.d.},\ t = 1,\ldots,T. \tag{6}$$

Conditioning on the initial observation y_0 (Phillips [70]) finds Jeffreys prior (or "critics' prior," in accordance with Phillips [71]) to be

$$\pi_{0,J}(\rho,\sigma) \propto \frac{1}{\sigma}\left(\alpha_0(\rho) + \frac{1-\rho^{2T}}{1-\rho^2}\left(\frac{y_0}{\sigma}\right)^2\right)^{1/2}, \tag{7}$$

where

$$\alpha_0(\rho) = \frac{1}{1-\rho^2}\left(T - \frac{1-\rho^{2T}}{1-\rho^2}\right), \tag{8}$$

and where $\pi_{0,J}(1,\sigma)$ and $\pi_{0,J}(-1,\sigma)$ are given by continuity. The critics' prior is defined for all $\rho \in \mathbf{R}$, $\sigma \geq 0$, is increasing in $\rho \geq 0$ and diverges quickly for $\rho > 1$.

Conditioning on $\sigma = 1$, Figure 1 compares the flat prior with the critics' prior with $T = 100$ and a few other priors described below. In Figure 1, ρ is restricted to the interval [0.5;1], and the priors shown all integrate to unity over that interval. Notice that the difference between the critics' prior and a flat prior is rather small.

OBSERVATION 5. *Conditional on nonexplosive roots* $|\rho| \leq 1$, *the differences between the critics' prior and a flat prior is small and will usually not matter in practical applications.*

The difference between a flat prior and the critics' prior becomes large once explosive roots are taken seriously. Figure 2 shows the same priors as Figure 1, but they are restricted to $\rho \in [0.6;1.1]$ rather then $\rho \in [0.5;1]$. Because the data can distinguish more easily between $\rho = 1.1$ and $\rho = 1.05$ than between $\rho = .75$ and $\rho = .7$, for example, the critics' prior assigns most of its weight to the explosive region $\rho > 1$.

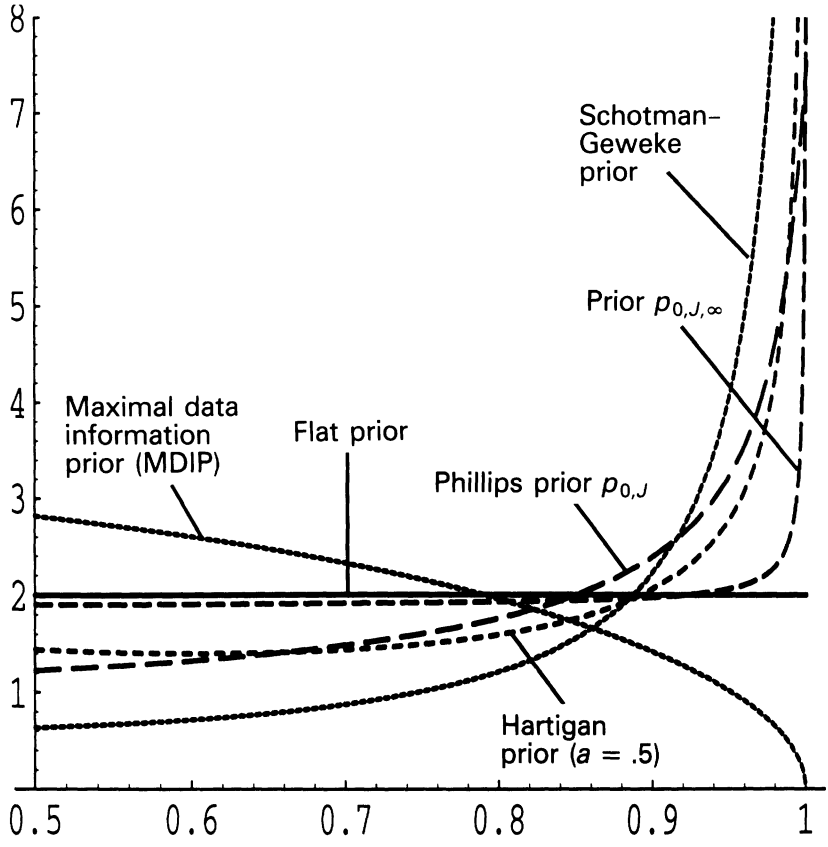

FIGURE 1. Comparing priors in the nonexplosive region $0.5 \leq \rho \leq 1$ for the model in equation (6). They do not differ much (Observation 5).

OBSERVATION 6. *The critics' prior exhibits an explosive behavior for explosive roots, i.e., roots above* 1 *in absolute value. To impose a rather flat or a declining prior for roots above* 1 *or a Normal-Wishart prior centered at the unit root or to set the prior to* 0 *beyond some* $\bar{\rho} \geq 1$ *corresponds to a prior belief that explosive roots are unlikely, and the more so, the more explosive they are.*

Most researchers probably wish to impose such a prior belief; see Koop and Steel [43], Leamer [55], Kim and Maddala [38], Poirier [79], Schotman and van Dijk [85], DeJong and Whiteman [17], and Sims [98], but also Phillips [71]. For example, DeJong and Whiteman [16–18] and DeJong [15] use a flat prior set to 0 outside the interval $\rho \in [0.55, 1.05]$.

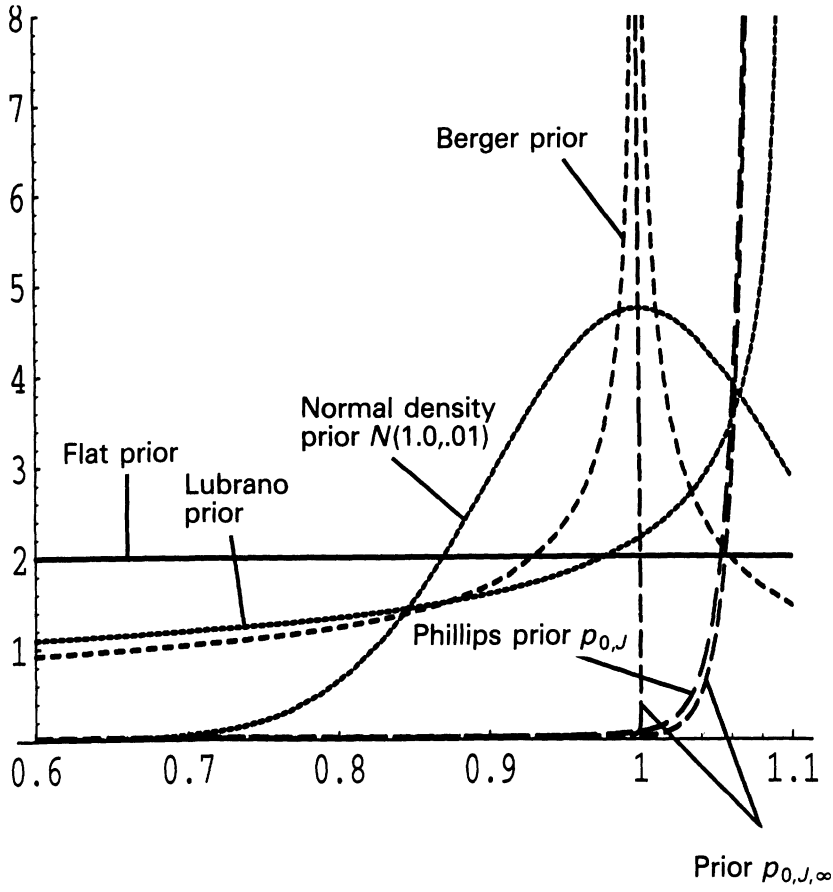

FIGURE 2. Comparing priors for a region including explosive roots, $0.6 \leq \rho \leq 1.1$ for the model in equation (6). The priors differ greatly in their treatment of the explosive region (Observation 6).

OBSERVATION 7. *The prior belief of excluding roots that are too explosive reflects a consensus belief of applied researchers and should therefore be employed as benchmark when doing and reporting Bayesian analysis of economic time series.*

Phillips's calculations were based on the conditional likelihood function rather than the exact likelihood function. By using the latter and extending results by Thornber [100] and Zellner [106], Uhlig [103] finds Jeffreys prior to be

$$\pi_{0,J,\infty}(\rho,\sigma) \propto \begin{cases} \dfrac{1}{\sigma^2}\exp\left(-\dfrac{y_0^2(1-\rho^2)}{2\sigma^2}\right)\left(\dfrac{4\rho^2}{1-\rho^2}+2(T+1)\right)^{1/2} & \text{for } |\rho|\le 1 \\ \dfrac{1}{\sigma^2}\left(2\dfrac{T+1}{T}\dfrac{1-\rho^{2T}}{1-\rho^2}\right)^{1/2} & \text{for } |\rho|>1. \end{cases} \tag{9}$$

The notation is as in equation (4). This prior has an integrable singularity[6] of order $(1-\rho)^{-1/2}$ at $\rho = 1$ (see Figures 1 and 2). By restricting this prior to the interval $\rho \in (-1;1)$, it converges pointwise to the flat prior σ^{-1} times the exact likelihood from the initial observation, as $T \to \infty$. Thus, the use of a flat prior $\pi(\rho,\sigma) \propto \sigma^{-1}$ for the nonexplosive region is justified because Jeffreys prior calculated from the exact likelihood function approximates it arbitrarily well for large samples (see Observation 8).

By using just the exact likelihood function for the initial observation to calculate Jeffreys prior, Thornber [100], Zellner[7] [106,107], and Lubrano [61] proposed using a Beta-distribution on some interval $\rho \in [-\bar\rho,\bar\rho]$ as prior π: Thornber and Zellner suggested $\bar\rho = 1$ and Lubrano $\bar\rho > 1$. Berger, in a seminar at Yale (1992), suggested extending the Jeffreys–Thornber–Zellner prior "symmetrically" through the transformation $\tilde\rho = 1/\rho$ to the explosive region, which results in the prior $\pi(\rho) = a(1-\rho^2)^{-1/2}$ for $|\rho| < 1$ and $\pi(\rho) = a(\rho^2-1)^{-1/2}|\rho|^{-1}$ for $|\rho| > 1$, where a is the appropriate integrating constant.

Sampling frequency may be an issue; see Sims [98]. Hartigan, in a comment at Yale (1992), suggested using a sampling-frequency invariant[8] prior $\pi_0(\rho) \propto 1/(\sigma\rho(-\log(\rho))^\alpha)$ for $0 \le \rho \le 1$, where α is a free parameter. Given σ, this prior has a nonintegrable singularity at[9] $\rho = 1$ for $\alpha \ge 1$, whereas it has a nonintegrable singularity at $\rho = 0$ for $\alpha \le 1$. Because these singularities will persist into the posterior, a reasonable choice for this prior may be to restrict ρ to the interval [0.5;1] and to choose $\alpha = .5$. Geweke and Schotman (see Schotman [88]) computed Jeffreys prior for a continuous time process $dy = \kappa y dt + \sigma dW$, $0 \le t \le T$, $y_0 = 0$ sampled at time intervals of h. As $h \to 0$, one obtains for $\rho = \exp(-\kappa)$,

$$\pi_0(\rho,\sigma) \propto \frac{1}{\sigma\rho(-\log(\rho))^{1/2}}\left(T - \frac{1-\rho^{2T}}{-2\log(\rho)}\right)^{1/2}, \tag{10}$$

where $0 \le \rho \le 1$ (note the similarity to the prior suggested by Hartigan). Given σ, the right-hand side of (10) converges to a finite limit for $\rho \to 1$ but has a nonintegrable singularity for $\rho \to 0$. It may be reasonable to restrict this prior to $\rho \in [0.5;1]$.

The MDIP for (6), by using the exact likelihood function, is given by $\pi(\rho,\sigma) \propto (1-\rho^2)^{1/2}/\sigma$ and converges to zero as $\rho \to 1$; see Zellner [106].

Overall, the convenient Normal–Wishart prior does not appear to be a bad choice for most applications when restricting ρ to $|\rho| \leq 1$.

OBSERVATION 8. *If a univariate autoregression with no time trend is to be analyzed, a Normal–Wishart prior centered around a random walk as prior π_0 is reasonable.*

Note, however, that some of the priors listed above produce singularities at the unit root itself: this can substantially affect the conclusions. Likewise, downweighing explosive roots a priori can have substantial impact on answers to some questions (see, e.g., the Bayesian unit root tests performed by computing the posterior probability $P(\rho \geq 1)$ in DeJong and Whiteman [16–18] and Phillips [70] or the predictive densities in Section 4). If feasible, one should therefore follow Leamer's [52,53] suggestion, echoed in Poirier [77], and analyze the sensitivity of the conclusions to the possible prior singularities at the unit root and the prior treatment of explosive roots if the latter are to be taken seriously.

3.2. Including a Trend

In most applications, a linear time trend needs to be included. Consider

$$y_t = \mu + \beta t + \rho y_{t-1} + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \sigma^2)\text{i.i.d.},\ t = 1, \ldots, T, \tag{11}$$

with the restriction that $\beta = 0$ when $\rho = 1$ to exclude a quadratic time trend when there is a unit root. Phillips [71, equation (8)] calculates that Jeffreys prior (conditional on the initial observation) is well approximated by

$$\pi_{0,P}(\mu, \beta, \rho, \sigma) \propto \frac{1}{\sigma^3}\, \alpha_0(\rho)^{1/2}. \tag{12}$$

A good way to understand (11) is to rewrite it as a components model

$$y_t = \gamma + \delta t + u_t \tag{13}$$

$$u_t = \rho u_{t-1} + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \sigma^2)\text{i.i.d.},\ t = 1, \ldots, T \tag{14}$$

(cf. DeJong et al. [19] and Schotman and van Dijk [85]). In this parameterization, the nature of the deterministic and the stochastic trends are nicely separated, whereas they are interdependent[10] in (11): because

$$\begin{aligned} \sigma &= \sigma \\ \rho &= \rho \\ \beta &= (1 - \rho)\delta \\ \mu &= \gamma(1 - \rho) + \rho\delta, \end{aligned} \tag{15}$$

a fixed value for β, for example, implies an ever steeper deterministic time trend as $\rho \to 1$.

For the components parameterization, Schotman and van Dijk [85] calculate Jeffreys prior $\pi_{0,J}$ by using the conditional likelihood and obtaining

$$\pi_{0,J}(\rho,\gamma,\delta,\sigma) \propto \frac{1}{\sigma^3}(1-\rho)^2\left(\alpha_0(\rho) + \frac{1-\rho^{2T}}{1-\rho^2}\left(\frac{y_0-\gamma}{\sigma}\right)^2\right)^{1/2}. \tag{16}$$

The corresponding prior $\pi_{0,J}(\rho,\beta,\mu,\sigma)$ for the reduced form (11) is obtained by multiplying with the determinant $(1-\rho)^{-2}$ of the Jacobian of the parameter transformation (15).[11] For $y_0 = \gamma$, one again obtains Phillips prior $\pi_{0,P}$ in equation (12).

By using the exact, rather than the conditional, likelihood function for the components specification in (13), Uhlig [103] calculated Jeffreys prior and found the corresponding prior $\pi_{0,J,\infty}$ for the "reduced form" specification in (11) through the transformation rule. Conditional on (μ,ρ,σ), this prior prescribes a particular normal distribution for β, the coefficient on the time trend:

$$\beta \sim \mathcal{N}\left(\frac{1-\rho}{\rho}\mu - \frac{(1-\rho)^2}{\rho}y_0, \frac{\sigma^2(1-\rho)^3}{\rho^2(1+\rho)}\right), \quad \text{if } |\rho| < 1. \tag{17}$$

The variance in this normal distribution degenerates to zero as $\rho \to 1$. This is a reasonable prior restriction given that β ought to be zero for $\rho = 1$ in (11). A practical approximation to (17) is

$$\beta \sim \mathcal{N}\left(0, \frac{\sigma^2}{2}(1-\rho)^3\right), \quad \text{if } |\rho| < 1, \tag{18}$$

conditional on (μ,ρ,σ). We modify the commonly used Normal-Wishart prior for (11) in this way. Further inspection of Jeffreys prior $\pi_{0,J,\infty}$ yields that this should be done by multiplying with the factor

$$f(\beta|\rho,\sigma) = \exp\left(\frac{-\beta^2}{\sigma^2(1-\rho)^3}\right), \quad |\rho| < 1 \tag{19}$$

rather than the full normal density function from (18).

Calculating the MDIP results in $\pi \propto (1-\rho^2)^{1/2}/\sigma$ for both specifications, (11) or (13). Thus, applying the transformation rule to the MDIP for (13) results in the prior $\pi \propto ((1+\rho)/(1-\rho))^{1/2}\sigma^{-1}$ for (11), which is quite different from the MDIP calculated directly for (11) because the MDIP is not immune to parameter transformations.

OBSERVATION 9. *If a univariate autoregression that includes a constant and a linear time trend is to be analyzed, if there is concern about at most one unit root or explosive root, and if a reduced-form model as in* (11) *is used, a Normal-Wishart prior as prior* π_0 *is reasonable, provided it is modified by multiplication with the factor f given by equation* (19) *in the region where* $|\rho| \leq 1$ (ρ *denotes the largest root*).[12]

If feasible, writing the model as a components model as in (13) is more natural and imposing a Normal–Wishart prior there is reasonable. For a computationally feasible extension of this approach, see Geweke [31].

3.3. Testing for a Unit Root

Once a prior is chosen, Bayesian tests of the unit root null hypothesis can be performed. DeJong and Whiteman [16–18] perform the test for

$$H_0: \rho \geq 1 \quad \text{vs.} \quad H_1: \rho < 1 \tag{20}$$

by calculating posterior masses for those regions using Normal–Wishart priors. Phillips [70] criticized these tests because they are sensitive to the prior treatment of explosive roots; see our earlier discussion. As Schotman and van Dijk [85] pointed out, the appropriate test for a unit root[13] often is

$$H_0: \rho = 1 \quad \text{vs.} \quad H_1: |\rho| < 1. \tag{21}$$

To perform these tests, Bayesian posterior odds ratios are the appropriate tool. They can be thought of as prior-averaged likelihood ratios. Care needs to be taken when some parameters are not identified under either alternative because that can result in nonsensical infinite integrals for certain improper priors. For details, see, e.g., Zellner [106], Schotman and van Dijk [85], and Koop [45].

4. PERSISTENCE AND FORECASTS

To provide a simple example for the key issue of persistence and medium-term forecasting, consider again the AR(1) process in (6), where just ρ is assumed to be unknown and where $\sigma^2 = 1.0$. Let $Y_T = \{y_0, \ldots, y_T\}$. Both the classical and the Bayesian econometricians agree that the impulse response of y_{T+n} to a shock ϵ_T of unit size is given by

$$r(n) = \rho^n. \tag{22}$$

With $\rho = 1$, a shock persists forever, whereas with $\rho = .8$ the halflife of a shock is just $n = 3.1$. Both groups also agree that a forecast is given by

$$\hat{y}_{T+n} \equiv E[y_{T+n} | \rho, Y_T] = \rho^n y_T \tag{23}$$

and

$$\text{Var}[y_{T+n} | \rho, Y_T] = \begin{cases} \sigma^2 \dfrac{1 - \rho^{2n}}{1 - \rho^2}, & \text{if } \rho \neq 1, \\ n\sigma^2, & \text{if } \rho = 1. \end{cases} \tag{24}$$

The Bayesian and the classicist differ however in their treatment of the uncertainty in ρ. We concentrate on the forecasting exercise because a persistence

analysis by impulse response functions follows a similar logic; compare (22) with (23).

Suppose that for some data set, $T = 10$, $y_0 = 0$, $y_T = 4$, $\hat{\rho} = 1.0$, and $\sigma_\rho^2 = .04$, where

$$\hat{\rho} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{25}$$

and

$$\sigma_\rho^2 = \frac{\sigma^2}{\sum_{t=1}^{T} y_{t-1}^2}. \tag{26}$$

This is a rather extreme situation chosen to demonstrate the differences in inference.

With these numbers, a classical econometrician would not reject the null hypothesis of a unit root. If this test is used as a pretest, the classicist would proceed by using the model

$$\Delta y_t = \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0,1)\text{i.i.d.},\ t = 1, \ldots, T. \tag{27}$$

For the forecasting exercise, a naive classicist may then conclude that

$$E[y_{T+n} | Y_T] = y_T \tag{28}$$

and

$$\text{Var}[y_{T+n} | Y_T] = n \tag{29}$$

for this particular sample. The uncertainty about ρ disappears in the forecasting step. The forecast distributions of the naive classicist for y_{T+n} are normal $\mathcal{N}(4, n)$ and are plotted in Figure 3. A naive classical econometrician ends up being too sure about his forecast: even though he does not actually know whether there is a unit root or not, these forecast distributions assume that he does and do not take the parameter uncertainty into account.

A sophisticated classical econometrician will surely object to this naive approach. He could, for example, construct unbiased estimators and confidence regions for the n-step-ahead mean forecast without pretesting for unit roots first.[14] Alternatively, he could evaluate the distribution of the n-step-ahead forecast by taking into account the pretesting procedure. Unfortunately, this is rarely done (see, e.g., Campbell and Perron [8], Cochrane [12], Fair [26], Sampson [82], and Elliot and Stock [25]). Even if such an analysis is performed, the resulting forecast distribution of the sophisticated classicist depends on the unknown parameter, which is not known when performing the forecast. It seems more natural to state the uncertainty about

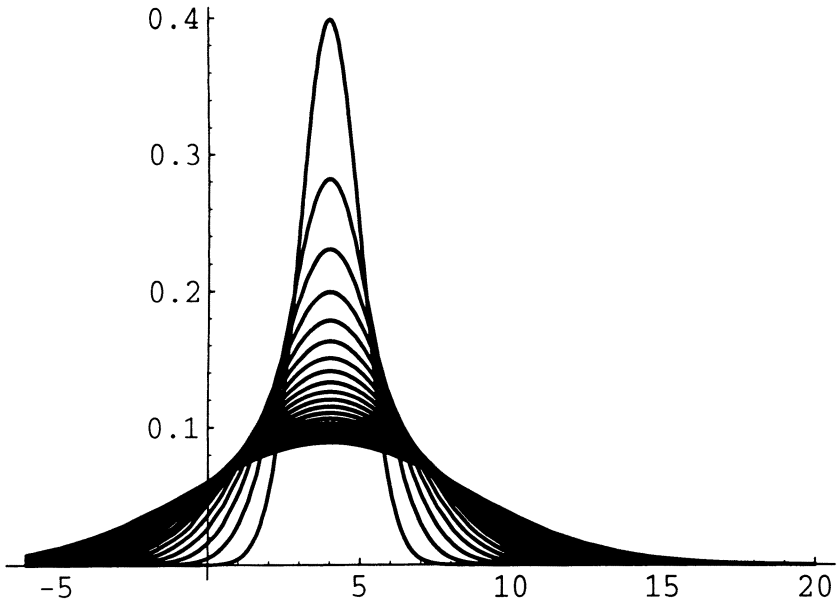

FIGURE 3. Classicists' n-step-ahead distribution of y_t.

a forecast by including parameter uncertainty and taking the observed data as given: this is what a Bayesian approach does.

A Bayesian econometrician naturally recognizes that there is remaining uncertainty about the parameter ρ. Consider first a flat prior over ρ. The posterior[15] is given by

$$\rho \sim \mathcal{N}(\hat{\rho}, \sigma_\rho^2). \tag{30}$$

As a result, there is uncertainty already about $\hat{y}_{T+n}$, the expected value of y_{T+n} given the data and ρ. The resulting densities for $\hat{y}_{T+n}$ or, alternatively, the uncertainty about $r(n)$, the value of the impulse response function at step n, can be calculated directly[16] and are plotted in Figure 4. The densities become increasingly skewed with a mode approaching zero and a long, fat, right tail as the forecasting horizon increases. The mode arises from the nonzero weight the posterior assigns to stationary roots $|\rho| < 1$, whereas the tail arises from the nonzero weight given to explosive roots. By adding the uncertainty about future ϵ_t, one obtains the predictive densities[17] for $y_{T+n}|Y_T$. They are plotted in Figure 5 for $n = 1, \ldots, 10$ steps ahead. They are not normal but skewed and quite different[18] from the densities in Figure 3. Clearly, the mass assigned by the posterior to the region $\rho \geq 1$ matters for evaluating the uncertainty about the n-step-ahead forecasts.

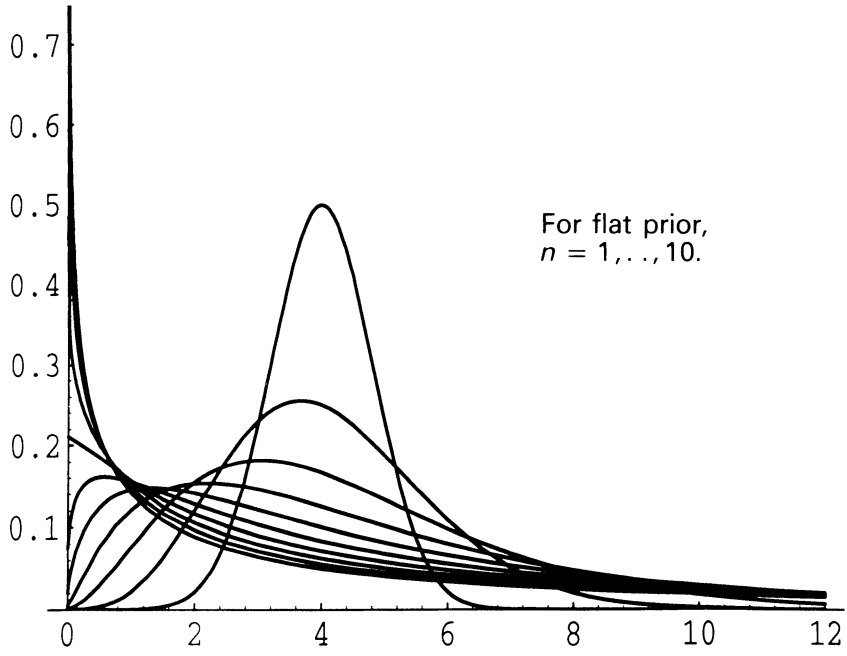

FIGURE 4. Bayesian n-step-ahead distribution of $E(y_t|r, \text{data})$.

Predictive densities that use any of the other priors discussed in the previous section can be created similarly: Figures 6 and 7 show five of them and the resulting posteriors together for comparison. Figures 8 and 9 show the predictive densities for y_t by using the critics' prior $\pi_{0,J}$ given in equation (7) as calculated by Phillips. In Figure 8, we restricted the prior (and thus the posterior) to $\rho \in [0;2.0]$, whereas we restricted the prior to $\rho \in [0;1.1]$ in Figure 9. In particular, the tails depend on this restriction. In Figure 10 the prior $\pi_{0,J,\infty}$ given in equation (9) is used to calculate the densities for the values of the impulse response function $r(n)$, $n = 1, \ldots, 10$. This figure is therefore similar in construction to Figure 4. There is a spike at the unit root that persists into the posterior. Figure 11 is similar to Figure 5 and shows the predictive density for y_{T+n}, when the prior $\pi_{0,J,\infty}$ is used and restricted to [0;2]. Figure 12 shows the same as Figure 11 except that the prior $\pi_{0,J,\infty}$ is now restricted to [0;1]. The singularity at $\rho = 1$ of the prior $\pi_{0,J,\infty}$ does not matter much when calculating the predictive densities rather than densities for the impulse response function values as in Figure 10, but the particular exclusion of explosive roots does.

OBSERVATION 10. *The predictive density and, in particular, its tails can be sensitive to the prior treatment of explosive roots. Sensitivity analysis should be performed if these roots are to be taken seriously.*

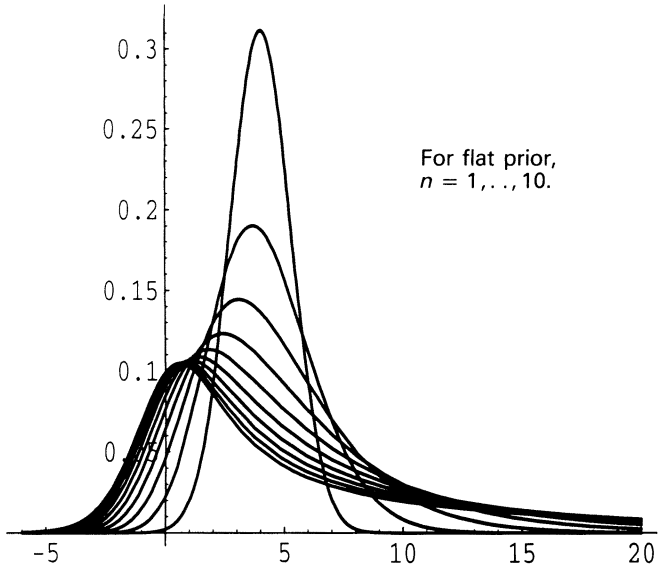

FIGURE 5. Predictive densities for y_t, n-steps ahead.

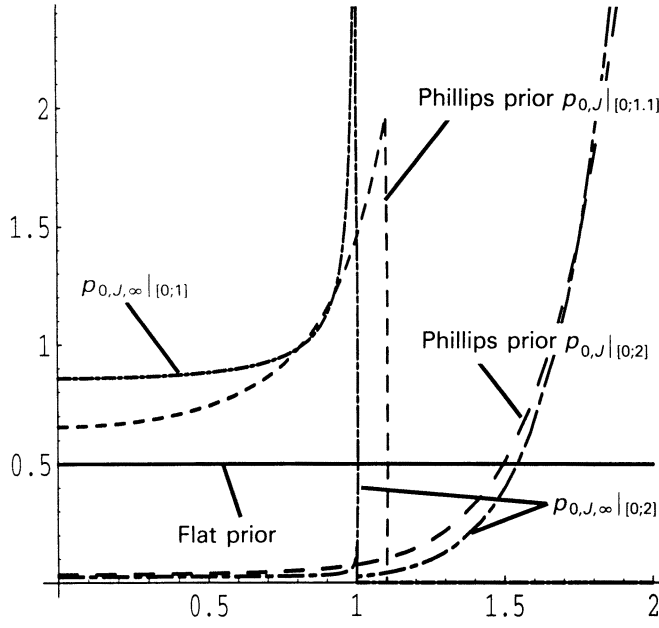

FIGURE 6. Priors.

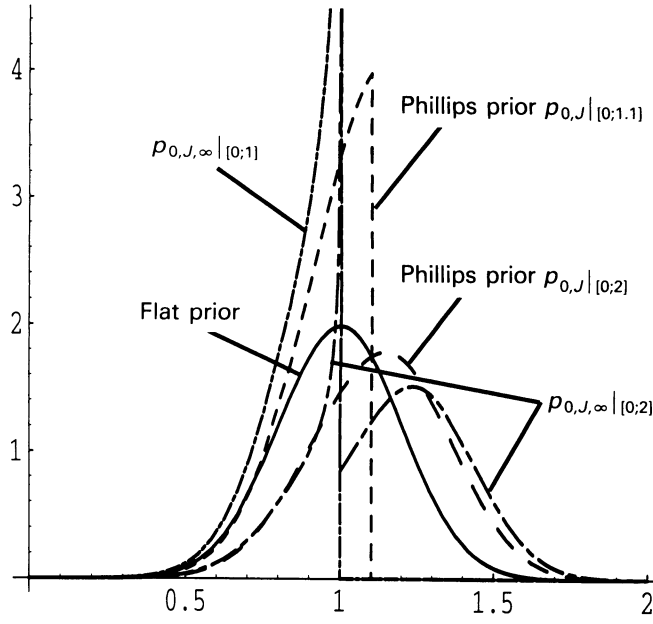

FIGURE 7. Posteriors to priors acc. to label.

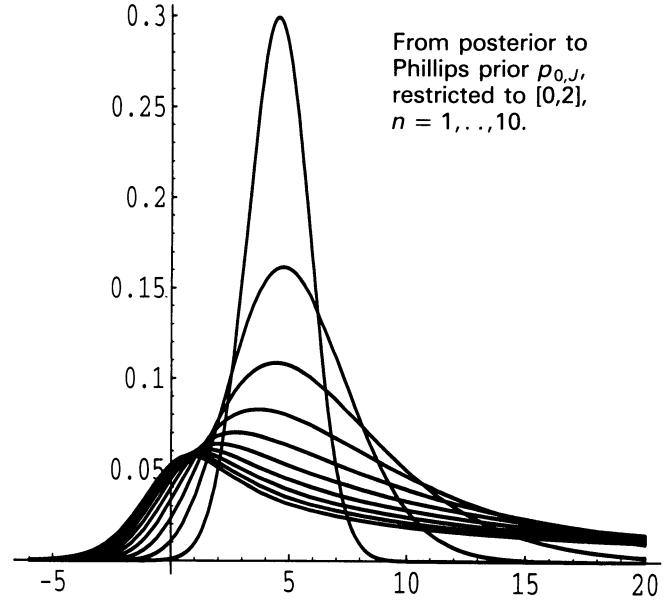

FIGURE 8. Predictive densities for y_t, n-steps ahead.

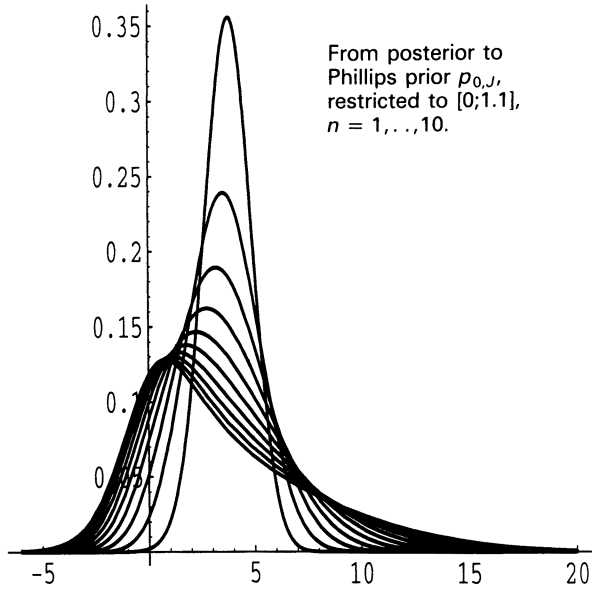

FIGURE 9. Predictive densities for y_t, n-steps ahead.

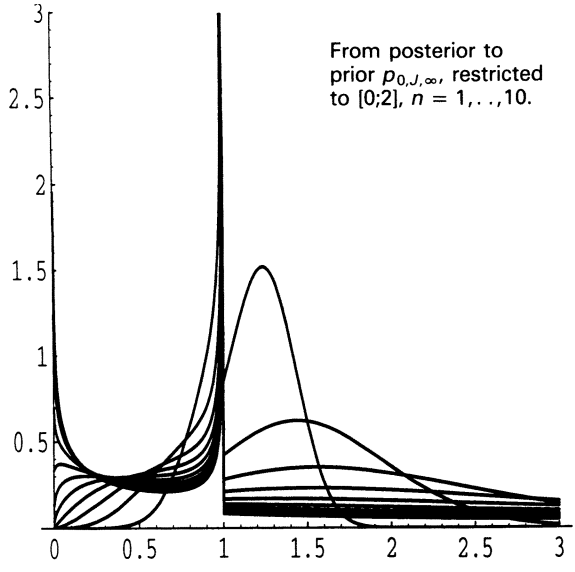

FIGURE 10. Densities for the value of the impulse response function $r(n)$, n-steps ahead.

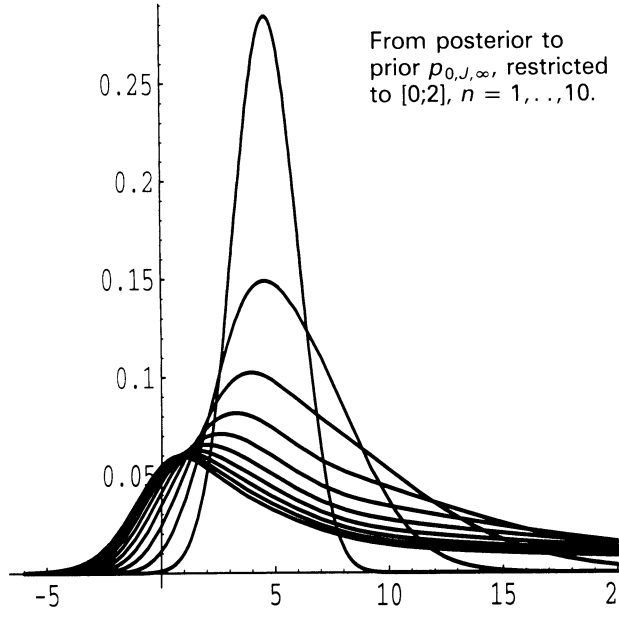

FIGURE 11. Predictive densities for y_t, n-steps ahead.

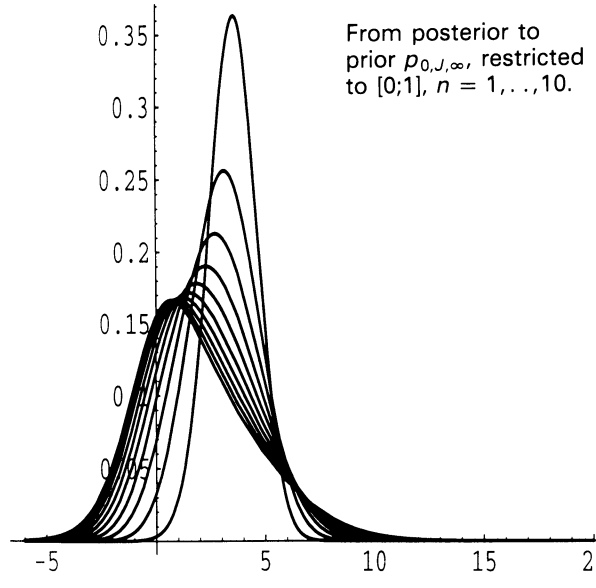

FIGURE 12. Predictive densities for y_t, n-steps ahead.

There is more uncertainty about the forecast in any of these figures except in Figure 3 because the Bayesian takes into account the uncertainty about the underlying coefficient, whereas a classicist may not when using unit root tests as a pretesting device.

OBSERVATION 11. *Pretesting for unit roots or trend stationarity and proceeding as if one is sure about the conclusion of this pretest can be misleading in calculating the uncertainty with regard to n-step-ahead forecasts, in particular, and in answering macroeconomic questions, in general.*

Recently, the predictive density has been reinterpreted in classical language as a "Bayes model" or a "Bayesian frame of reference" and used to derive classical tests for unit roots (see Phillips and Ploberger [72,73]), model selection criteria, and prediction (see Phillips [74–76]).

The Bayesian approach to possibly nonstationary time series has proven fruitful in applications as evidenced by the work of Doan, Litterman, and Sims [22], Sims [91,93,95], DeJong and Whiteman [15,16,18,21], Geweke [31], Koop [45,46], McCulloch and Tsay [63], Phillips [74–76], Schotman and van Dijk [83,84,86], Tsurumi and Wago [101], Zellner, Hong, and Gulati [109], Zellner [111], Zivot and Phillips [113], and Zivot [114].

5. CONCLUSIONS

In this paper, we summarized some of the recent literature on the Bayesian treatment of unit roots into eleven observations for applied macroeconomists. Macroeconomists are typically interested in some persistence property or medium-run forecasting property. Therefore, Bayesian methods are especially suitable.

In Section 2, we reviewed Bayesian methodology for time series analysis and discussed the choice of priors. We summarized and extended the recent discussion about appropriate choices of priors. By recognizing a consensus distaste for explosive roots, the popular Normal–Wishart priors centered at the unit root are found to be reasonable, provided they are modified in reduced-form models by concentrating the prior mass for the time trend coefficient toward zero as the largest root approaches unity from below.

As for persistence and medium-term forecasting, Bayesian methods naturally take the uncertainty about the presence of a unit root into account. The tails of the predictive densities can be sensitive to the prior treatment of explosive roots.

NOTES

1. For this reason, some endorse comparing the relative sizes of the temporary and the permanent components of a shock rather than testing for the existence of a persistent component per se (see, e.g., Watson [104], Campbell and Mankiw [7], and Quah [80]).

2. Bayesian analysis is not restricted to normally distributed errors. For other types of distributions or nonconjugate priors, numerical integration methods as in, e.g., Kloek and van Dijk [42], Naylor and Smith [65], and Geweke [27,29,30] are available. If the error distribution is unspecified, Bayesian limited information methods can be used; see, e.g., Kwan [49] and Kim [39].

3. See Appendix for a definition of a Normal-Wishart distribution.

4. See Appendix for a definition of a Normal-Wishart distribution.

5. For the beginning debate about appropriate priors for multivariate models, see, e.g., Kleibergen and van Dijk [41] and DeJong and Whiteman [20].

6. A function $f(x)$ has a singularity at $\bar{x}$ if $f(x)$ diverges as $x \to \bar{x}$. The singularity is integrable if $\int_N |f(x)|\,dx < \infty$ for some open neighbourhood N of $\bar{x}$, and nonintegrable otherwise.

7. Zellner investigates (6) by including a constant; his calculations are easily modified, however.

8. A prior π_0 for equation (6) is sampling-frequency invariant if it is invariant under the transformation $\rho \mapsto \rho^k$, i.e., if $\pi_0(\rho) \propto \pi_0(\rho^k)\rho^{k-1}$ for any $k > 0$.

9. This is easy to see with the parameter transformation $\kappa = -\log(\rho)$.

10. See Schotman [87] for a further discussion of these issues.

11. Strictly speaking, the transformation rule is only valid for $\rho \neq 1$. Because we are interested mostly in the shape of priors rather than their exact value at $\rho = 1$, we ignore this as a technicality.

12. To calculate the posterior, use just the Normal-Wishart prior with the formulas in (A.5)–(A.8). The resulting posterior must then be multiplied with the factor f from equation (19) in the region where $|\rho| \leq 1$ and renormalized to integrate to 1.

13. A continuous prior will assign probability zero to any nullset and therefore probability zero to the region characterized by the unit root $\rho = 1$ in (11). However, by using a Beveridge-Nelson MA representation instead, $y_t = \gamma + \delta t + \alpha W_t + A(L)\epsilon_t$, where $W_t = \sum_{j=0}^{t} \epsilon_j$ and $A(L)$ is stationary, a continuous posterior over the coefficients γ, δ, α and $A(L)$ will assign probability zero to the trend-stationary subset of the parameter space because it is characterized by $\alpha = 0$. The reason for this difference is restrictions on the spectral density for Δy_t introduced by the particular parameterization. It may be interesting to use the Bayesian method directly for inference about the mass of the spectral density of Δy_t in some persistence-relevant region $[0, \nu]$, ν small.

14. In particular, Andrews [1] argued for correcting the LSE so as to obtain a median-unbiased estimator $\tilde{\rho}$ of ρ rather than a mean-unbiased estimator. The median-unbiased impulse response function is then simply given by $\tilde{r}(n) = \tilde{\rho}^n$.

15. If the prior is fixed and continuous, or at least converges to a continuous proper prior in the limit as the sample size increases to infinity, the posterior will, in fact, be asymptotically normal, as shown by Kim [40]; for a general treatment, see also Phillips and Ploberger [72]. Thus a normal-shaped posterior is a good benchmark from which to analyze the forecasting issue.

16. Given the posterior π_T, the density for the mean forecast is given by

$$f(y) = \frac{1}{n y_T}\,\pi_T\left(\left(\frac{y}{y_T}\right)^{1/n}\right)\left(\frac{y}{y_T}\right)^{1/n-1},$$

ignoring some difficulties that arise from negative ρ. The density for $r(n)$, the value of the impulse response function at step n, is given by substituting $y_T = 1$ in this expression.

17. For general results regarding these predictive densities, see Chow [10], who has explicitly calculated the mean for multivariate autoregressive model predictions if a Normal-Wishart prior is used; see also Koop, Osiewalski, and Steel [47,48].

18. Given any posterior, these distributions converge for $n \to \infty$ to a limit that is given by the "weighted sum" of all stationary distributions $F_\rho(y)$ for y drawn from (6) with $|\rho| < 1$, where the weight for F_ρ is given by the posterior height at ρ, as well as a mass concentrated

at $\pm\infty$ equal to the mass assigned by the posterior to explosive roots $|\rho| \geq 1$. The median of these distributions is given by $\tilde{\rho}^n y_T$, where $\tilde{\rho}$ is the median of the posterior: this is analogous to Andrews's [1] analysis.

REFERENCES

1. Andrews, D.W.K. Exactly unbiased estimation of first order autoregressive/unit root models. *Econometrica* 61 (1993): 139–166.
2. Box, G.E.P. & G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison Wesley, 1973.
3. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag, 1985.
4. Berger, J.O. & R.L. Wolpert. *The Likelihood Principle*. Hayward, CA: Institute of Mathematical Statistics, 1988.
5. Broemeling, L.D. *Bayesian Analysis of Linear Models*. New York: Marcel Dekker, 1985.
6. Broemeling, L.D. & H. Tsurumi. *Econometrics and Structural Change*. New York: Marcel Dekker, 1987.
7. Campbell, J.Y. & N.G. Mankiw. Are output fluctuations transitory? *Quarterly Journal of Economics* 102 (1987): 857–880.
8. Campbell, J.Y. & P. Perron. Pitfalls and opportunities: What macroeconomists should know about unit roots. In O.J. Blanchard & S. Fisher (eds.), *NBER Macroeconomics Annual*. Cambridge: MIT Press, 1991.
9. Canova, F. *Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting*. Department of Economics, European University Institute, 1991.
10. Chow, G.C. Multiperiod predictions from stochastic difference equations by Bayesian methods. *Econometrica* 41 (1973): 109–118.
11. Christiano, L.J. & M. Eichenbaum. Unit roots in GNP: Do we know, and do we care? *Carnegie-Rochester Conference Series on Public Policy* 32 (1990): 7–62.
12. Cochrane, J.H. How big is the random walk in GNP? *Journal of Political Economy* 96 (1988): 893–920.
13. Cochrane, J.H. A critique of the application of unit root tests. *Journal of Economic Dynamics and Control* 15 (1991): 275–284.
14. Cochrane, J.H. Comment. In O.J. Blanchard & S. Fisher (eds.), *NBER Macroeconomics Annual*. Cambridge: MIT Press, 1991.
15. DeJong, D.N. Co-integration and trend-stationarity in macroeconomic time series: Evidence from the likelihood function. *Journal of Econometrics* 52 (1992): 347–370.
16. DeJong, D.N. & C.H. Whiteman. The temporal stability of dividends and stock prices: Evidence from the likelihood function. *American Economic Review*, 81 (1991): 600–617.
17. DeJong, D.N. & C.H. Whiteman. The case for trend-stationarity is stronger than we thought. *Journal of Applied Econometrics*, 6 (1991): 413–422.
18. DeJong, D.N. & C.H. Whiteman. Reconsidering trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 28 (1991): 221–254.
19. DeJong, D.N., J.C. Nankervis, N.E. Savin & C.H. Whiteman. Integration versus trend stationarity in time series. *Econometrica* 60 (1992): 423–433.
20. DeJong, D.N. & C.H. Whiteman. Modeling stock prices without pretending to know how to induce stationarity. *Econometric Theory* 10 (1994): 701–719.
21. DeJong, D.N. & C.H. Whiteman. *The Forecasting Attributes of Trend- and Difference-Stationary Representations for Macroeconomic Time Series*. University of Pittsburgh, 1993.
22. Doan, T., R. Litterman & C. Sims. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3 (1984): 1–100.
23. Doan, T. *RATS Manual*. Evanston, IL: VAR Econometrics, 1990.

24. Durlauf, S.N. Output persistence, economic structure, and the choice of stabilization policy. *Brookings Papers on Economic Activity* 2 (1989): 69–136.
25. Elliot, G. & J.H. Stock. *Inference in Time Series Regressions When There Is Uncertainty about Whether a Regressor Contains a Unit Root*. Harvard University, 1992.
26. Fair, R.C. Estimating Event Probabilities from Macroeconometric Models Using Stochastic Simulation. NBER technical working paper #111, 1991.
27. Geweke, J. Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics* 38 (1988): 73–90.
28. Geweke, J. Comment of Poirier: Operational Bayesian methods in econometrics. *Journal of Economic Perspectives* 2 (1988): 159–166.
29. Geweke, J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57 (1989): 1317–1339.
30. Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Research Department Staff Report 148, Federal Reserve Bank of Minneapolis, 1991.
31. Geweke, J. Priors for macroeconomic time series and their application. *Econometric Theory* 10 (1994): 609–632.
32. Hendry, D.F. & J.-F. Richard. On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics* 20 (1982): 3–33.
33. Hendry, D.F. Econometric methodology: A personal perspective. In T. Bewley (ed.), *Advances in Econometrics, vol. 2*, pp. 29–48. Cambridge: Cambridge University Press, 1987.
34. Highfield, R.A. *Forecasting with Bayesian State Space Models*. Ph.D. Dissertation, University of Chicago, 1986.
35. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London A* 186 (1946): 453–461.
36. Jeffreys, H. *Theory of Probability*, 3rd ed. London: Oxford University Press, 1961.
37. Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl & T.-C. Lee. *The Theory and Practice of Econometrics*, 2nd ed. New York: John Wiley and Sons, 1985.
38. Kim, I.-M. & G.S. Maddala. Flat priors vs. ignorance priors in the analysis of the AR(1) model. *Journal of Applied Econometrics* 6 (1991): 375–380.
39. Kim, J.-Y. *A Limited Information Bayesian Procedure: I. Theoretical Investigation*. University of Minnesota, 1991.
40. Kim, J.-Y. *Large Sample Properties of Posterior Densities in a Unit-Root Time Series Model*. University of Minnesota, 1991.
41. Kleibergen, F. & H.K. van Dijk. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* 10 (1994): 514–551.
42. Kloek, T. & H.K. van Dijk. Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. *Econometrica* 46 (1978): 1–20.
43. Koop, G. & M.F.J. Steel. A comment on "To criticize the critics: An objective Bayesian analysis of stochastic trends," by P.C.B. Phillips. *Journal of Applied Econometrics* 6 (1991): 365–370.
44. Koop, G. & M.F.J. Steel. A Decision Theoretic Analysis of the Unit Root Hypothesis Using Mixtures of Elliptical Models. Discussion paper 9150, Center for Economic Research, Tilburg University, 1991.
45. Koop, G. Objective Bayesian unit root tests. *Journal of Applied Econometrics* 7 (1992): 65–82.
46. Koop, G. Aggregate shocks and macroeconomic fluctuations: A Bayesian approach. *Journal of Applied Econometrics* 7 (1992): 395–411.
47. Koop, G., J. Osiewalski & M.F.J. Steel. *Bayesian Long-Run Prediction in Time-Series Models*. Boston University, 1992.
48. Koop, G., J. Osiewalski & M.F.J. Steel. Posterior Inference on Long-Run Impulse Responses. Working paper 92-29, División de Economía, Universidad Carlos III de Madrid, 1992.

49. Kwan, Y.-K. *Bayesian Analysis with an Unknown Likelihood Function—A Limited Information Approach*. University of Chicago, 1990.
50. Kydland, F.E. & E.C. Prescott. The Econometrics of the General Equilibrium Approach to Business Cycles. Staff Report 130, Federal Reserve Bank of Minneapolis, 1990.
51. Leamer, E.E. *Specification Searches*. New York: John Wiley & Sons, 1978.
52. Leamer, E.E. Let's take the con out of econometrics. *American Economic Review* 73 (1983): 31–43.
53. Leamer, E.E. Sensitivity analysis would help. *American Economic Review* 75 (1983): 308–313.
54. Leamer, E.E. Econometric metaphors. In T. Bewley (ed.), *Advances in Econometrics, vol. 2*, pp. 1–28. Cambridge: Cambridge University Press, 1987.
55. Leamer, E.E. Comment on "To criticize the critics." *Journal of Applied Econometrics* 6 (1991): 371–374.
56. Litterman, R. Techniques of Forecasting Using Vector Autoregressions. Working paper 115, Federal Reserve Bank of Minneapolis, 1979.
57. Litterman, R. A Bayesian Procedure for Forecasting with Vector Autoregressions. Working paper, Massachusetts Institute of Technology, 1980.
58. Litterman, R. Forecasting and policy analysis with Bayesian vector autoregression models. *Federal Reserve Bank of Minneapolis Quarterly Review* (1984): 30–41.
59. Litterman, R. Specifying Vector Autoregressions for Macroeconomic Forecasting. Staff Report 92, Federal Reserve Bank of Minneapolis, 1984.
60. Litterman, R. Forecasting with Bayesian vector autoregressions—Five years of experience. *Journal of Business and Economic Statistics* 4 (1986): 25–38.
61. Lubrano, M. *Testing for Unit Roots and Cointegration in a Bayesian Framework*. CREQE, Marseille, 1991.
62. Lütkepohl, H. *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag, 1990.
63. McCulloch, R.E. & R.S. Tsay. Bayesian inference of trend- and difference-stationarity. *Econometric Theory* 10 (1994): 596–608.
64. Muirhead, R.J. *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons, 1982.
65. Naylor, J.C. & A.F.M. Smith. Econometric illustrations of novel numerical integration strategies for Bayesian inference. *Journal of Econometrics* 38 (1988): 103–125.
66. Pagan, A. Three econometric methodologies: A critical appraisal. *Journal of Econometric Surveys* 1 (1987): 3–23.
67. Pagan, A. Comment on Poirier: Dogma or doubt? *Journal of Economic Perspectives* 2 (1988): 153–158.
68. Pagan, A.R. & M.R. Wickens. A survey of some recent econometric methods. *Economic Journal* 99 (1989): 962–1025.
69. Phillips, P.C.B. Reflections on econometric methodology. *Economic Record* 64 (1988): 544–559.
70. Phillips, P.C.B. To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics* 6 (1991): 333–364.
71. Phillips, P.C.B. Bayesian routes and unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics* 6 (1991): 3435–3474.
72. Phillips, P.C.B. & W. Ploberger. *Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations*. Cowles Foundation for Research in Economics, Yale University, 1991.
73. Phillips, P.C.B. & W. Ploberger. Posterior odds testing for a unit root with data-based model selection. *Econometric Theory* 10 (1994): 774–808.
74. Phillips, P.C.B. *Bayes Models and Forecasts of Australian Macroeconomic Time Series*. Cowles Foundation for Research in Economics, Yale University, 1992.
75. Phillips, P.C.B. *Bayesian Model Selection and Prediction with Empirical Applications*. Cowles Foundation for Research in Economics, Yale University, 1992.

76. Phillips, P.C.B. *Bayes Methods for Trending Multiple Time Series with an Empirical Application to the U.S. Economy*. Cowles Foundation for Research in Economics, Yale University, 1992.
77. Poirier, D.J. Frequentist and subjectivist perspectives on the problems of model building in economics. *Journal of Economic Perspectives* 2 (1988): 121–144.
78. Poirier, D.J. The subjectivist response. *Journal of Economic Perspectives* 2 (1988): 167–170.
79. Poirier, D.J. A comment on "To criticize the critics: An objective Bayesian analysis of stochastic trends." *Journal of Applied Econometrics* 6 (1991): 381–386.
80. Quah, D. The relative importance of permanent and transitory components: Identification and some theoretical bounds. *Econometrica* 60 (1992): 107–118.
81. Rust, J. Comment on Poirier: The subjective perspective of a "Spiritual Bayesian." *Journal of Economic Perspectives* 2 (1988): 145–151.
82. Sampson, M. The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics* 6 (1991): 67–76.
83. Schotman, P.C. & H.K. van Dijk. A Bayesian analysis of the unit root in real exchange rates. *Journal of Econometrics* 49 (1991): 195–238.
84. Schotman, P.C. & H.K. van Dijk. *Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP*. Econometric Institute, Erasmus University, 1990.
85. Schotman, P.C. & H.K. van Dijk. On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6 (1991): 387–402.
86. Schotman, P.C. When Unit Roots Matter: Excess Volatility and Excess Smoothness of Long Term Interest Rates. Discussion paper 44, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1991 (rev. 1992).
87. Schotman, P.C. *Objective Information and Ignorance: Notes on Some Priors in the Bayesian Analysis of the AR(1) Model*. University of Limburg, 1992.
88. Schotman, P.C. *Jeffreys' Prior in Continuous Time*. Princeton University, 1992.
89. Shannon, C.E. The mathematical theory of communication. *Bell System Technical Journal* (1948).
90. Sims, C.A. The role of approximate prior restrictions in distributed lag estimation. *Journal of the American Statistical Association* 67 (1972): 169–175.
91. Sims, C.A. Are forecasting models usable for policy analysis? *Federal Reserve Bank of Minneapolis Quarterly Review* (1986): 2–15.
92. Sims, C.A. Making economics credible. In T. Bewley (ed.), *Advances in Econometrics, vol. 2*, pp. 49–60. Cambridge: Cambridge University Press, 1987.
93. Sims, C.A. Identifying Policy Effects. Working paper 351, Federal Reserve Bank of Minneapolis, 1987.
94. Sims, C.A. Bayesian skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12 (1988): 463–474.
95. Sims, C.A. A Nine Variable Probabilistic Macroeconomic Forecasting Model. Discussion paper 14, Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, 1989.
96. Sims, C.A. Modelling Trends. Discussion paper, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1990.
97. Sims, C.A. & H. Uhlig. Understanding unit rooters: A helicopter tour. *Econometrica* 59 (1991): 1591–1600.
98. Sims, C.A. Comment by Christopher A. Sims on "To criticize the critics: An objective Bayesian analysis of stochastic trends," by P.C.B. Phillips. *Journal of Applied Econometrics* 6 (1991): 423–434.
99. Stock, J.H. Bayesian approaches to the "unit root" problem: A comment. *Journal of Applied Econometrics* 6 (1991): 413–422.
100. Thornber, H. Finite sample Monte Carlo studies: An autoregressive illustration. *Journal of the American Statistical Association* 62 (1967): 801–818.

101. Tsurumi, H. & H. Wago. *Bayesian Inference of Unit Root Under an Unknown Order of an ARM(p,q) Error Process and an Application to the Yen Foreign Exchange Rate*. Rutgers University, 1992.
102. Uhlig, H. *Bayesian Vectorautoregressions with Stochastic Volatility*. Princeton University, 1991.
103. Uhlig, H. On Jeffreys prior when using the exact likelihood function. *Econometric Theory* 10 (1994): 633–644.
104. Watson, M.W. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18 (1986): 49–75.
105. West, M. & J. Harrison. *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 1989.
106. Zellner, A. *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons, 1971.
107. Zellner, A. Maximal data information prior distributions. In A. Aykac & C. Brumat (eds.), *New Developments in the Applications of Bayesian Methods*, pp. 211–232. Amsterdam: North Holland, 1977.
108. Zellner, A. Bayesian analysis in econometrics. *Journal of Econometrics* 37 (1988): 27–50.
109. Zellner, A., C. Hong & G.M. Gulati. Turning points in economic time series, loss structures and Bayesian forecasting. In S. Geisser, J.S. Hodges, S.J. Press & A. Zellner (eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics*, pp. 371–393. Amsterdam: North-Holland, 1990.
110. Zellner, A. Bayesian methods and entropy in economics and econometrics. In W.T. Grandy, Jr. & L.H. Schick (eds.), *Maximum Entropy and Bayesian Methods*, pp. 17–31. Kluwer Academic Publishers, 1991.
111. Zellner, A. Time Series Analysis, Forecasting and Econometric Modeling: The Structural Econometric Modeling, Time Series Analysis (SEMTSA) Approach. University of Chicago, 1991.
112. Zellner, A. & C. Min. Bayesian Analysis, Model Selection and Prediction. University of Chicago, 1992.
113. Zivot, E. & P.C.B. Phillips. A Bayesian Analysis of Trend Determination in Economic Time Series. Discussion paper 1002, Cowles Foundation, Yale University, 1991.
114. Zivot, E. A Bayesian analysis of the unit root hypothesis within an unobserved components model. *Econometric Theory* 10 (1994): 552–578.

APPENDIX: THE LIKELIHOOD FUNCTION AND THE NORMAL–WISHART PRIOR

Let $l \geq 1$ be an integer. Let N be $l \times l$ and positive definite, let S be $m \times m$ and positive definite, let $\bar{B}$ be an $l \times m$-dimensional matrix, and let $\nu \geq m$. The Normal–Wishart distribution $\phi_{NW}(B,H|\bar{B},N,S,\nu)$ for an $l \times m$-dimensional matrix B and an $m \times m$ matrix H is given by the density

$$\phi_{NW}(B,H|\bar{B},N,S,\nu) = \kappa(N,S,\nu,m)|H|^{l/2}\exp\left(-\tfrac{1}{2}\mathrm{vec}(B-\bar{B})'[H\otimes N]\mathrm{vec}(B-\bar{B})\right)$$
$$\times |H|^{(\nu-m-1)/2}\exp(-\tfrac{1}{2}\nu trHS),$$

where $\kappa(N,S,\nu,m)$ is the appropriate integrating constant. The Normal–Wishart distribution specifies that the precision H follows a Wishart distribution $\mathcal{W}_m(S^{-1}/\nu,\nu)$

with mean $E[H] = S^{-1}$ and that conditional on H, the matrix B in its columnwise vectorized form vec(B) follows a Normal distribution $\mathcal{N}(\text{vec}(\bar{B}), H^{-1} \otimes N^{-1})$. More information about the Wishart distribution can be found in Muirhead [64] or Zellner [106].

To analyze the likelihood function for the model in (1) and (2), rewrite it as

$$\mathbf{Y} = \mathbf{XB} + \epsilon, \tag{A.1}$$

where $X_t = [Y'_{t-1} Y'_{t-2} \cdots Y'_{t-k} 1 t]'$, $\mathbf{Y} = [Y_1 \cdots Y_T]'$, $\mathbf{X} = [X_1 \cdots X_T]'$, and $\epsilon = [\epsilon_1 \cdots \epsilon_T]'$. Conditional on the initial observations Y_t, $t = 1-k, \ldots, 0$, the likelihood function in B and the precision matrix $H = \Sigma^{-1}$ is

$$L(B,H|\mathbf{Y}) = (2\pi)^{-mT/2} |H|^{T/2} \exp\left(-\tfrac{1}{2}(\beta - \hat{\beta})'[H \otimes \mathbf{X}'\mathbf{X}](\beta - \hat{\beta})\right)$$

$$\exp\left(-\frac{T}{2} trHS\right), \tag{A.2}$$

where

$$\beta = \text{vec}(B), \qquad \hat{\beta} = \text{vec}(\hat{B}) \tag{A.3}$$

(with vec$(\cdot)$ denoting columnwise vectorization) and where

$$\hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \; \hat{\Sigma} = \frac{1}{T}(\mathbf{Y} - \mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B}) \tag{A.4}$$

are the MLE's (see, e.g., Zellner [106, Sect. 8.1]). Thus, a function in B and H, the likelihood function is proportional to a Normal-Wishart density $\phi_{NW}(B,H|\hat{B},N,(T/\nu)\hat{\Sigma},\nu)$ with $l = km + 2$ and $\nu = T - (k-1)m - 1$.

PROPOSITION 1. *If the prior π_0 is given by a Normal-Wishart density*

$$\phi_{NW}(B,H|B_0,N_0,S_0,\nu_0),$$

then the posterior π_T is given by a Normal-Wishart density

$$\phi_{NW}(B,H|B_T,N_T,S_T,\nu_T)$$

as well, where

$$\nu_T = T + \nu_0, \tag{A.5}$$

$$N_T = N_0 + \mathbf{X}'\mathbf{X}, \tag{A.6}$$

$$\bar{B}_T = N_T^{-1}(N_0 B_0 + \mathbf{X}'\mathbf{X}\hat{B}) \tag{A.7}$$

$$S_T = \frac{\nu_0}{\nu_T} S_0 + \frac{T}{\nu_T}\hat{\Sigma} + \frac{1}{\nu_T}(\hat{B} - \bar{B}_0)'N_0 N_T^{-1}\mathbf{X}'\mathbf{X}(\hat{B} - \bar{B}_0). \tag{A.8}$$

Proof. Adapt the proof of Leamer [51, Theorem 3.9]. ■

A special case is the "flat prior" (see, e.g., DeJong [15], Broemeling [5], Broemeling and Tsurumi [6], and Koop [46]): $\bar{B}_0$ is chosen arbitrarily, $N_0 = 0$, S_0 is chosen arbitrarily and $\nu_0 = -l$. (It is my opinion that $\nu_0 = 0$ is a more logical choice for a "flat" prior.) The "flat" prior is simply proportional to $|H|^{-(m+1)/2}$, and one obtains a Normal-Wishart posterior with $\nu_T = T - l$, $N_T = \mathbf{X}'\mathbf{X}$, $\bar{B}_T = \hat{B}$, and $S_T = \hat{\Sigma}$.

Alternatively, a Normal density prior for vec(B), which is a Normal–Wishart prior with a fixed covariance matrix $\Sigma = S_0 = \bar{\Sigma}$ and $\nu_0 \to \infty$, is popular in practice. It is commonly centered around the random-walk mean and lagged variables are "softly" excluded by tightening the prior around the value of zero for their coefficients; see, e.g., Sims [90,91,93,95], Litterman [56–60], Doan, Litterman, and Sims [22], Highfield [34], the RATS manual [23], Lütkepohl [62], Canova [9], and Uhlig [102]. This prior is sometimes referred to as the "Minnesota prior." The software RATS [23] performs the required calculations.

One can understand Proposition 1 as a proposition about signal extraction, when conditioning on H: the prior corresponds to some imprecise, normally distributed signal about the unknown parameters and so does the data. When combining these two, one obtains a posterior corresponding to just one signal centered at a weighted sum of the centers with a precision equal to the sum of the precisions. In applications it is sensible to choose priors that are much more "imprecise" than the data about the important aspects of the parameters.