WHAT MACROECONOMISTS SHOULD KNOW ABOUT UNIT ROOTS AS WELL: THE BAYESIAN PERSPECTIVE*

Harald Uhlig
Princeton University

Econometric Research Program
Research Memorandum No. 367

May 1991
Revised June 1992

*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, Herman van Dijk, Pierre Perron, Peter Phillips, Albrecht Ritschl, Peter Schotman and James Stock.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, USA

Abstract

This paper summarizes recent Bayesian research on unit roots for the applied macroeconomist in the way Campbell and Perron [9] have recently summarized the classical unit roots perspective. The approach to the unit root problem should depend on the focus of a particular investigation, which I argue to be typically some persistence property or medium-term forecasting property. Bayesian methods are therefore especially suitable. The appropriate choice of a prior is discussed. 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 towards zero as the largest root approaches unity from below. I discuss that the tails of the predictive density can be sensitive to the prior treatment of explosive roots. I conclude that Bayesian methods often deliver more natural answers to macroeconomic questions than classical methods.

Bayesians are like the Hare-Krishna types: you get too close and you might become one of them. (Jon Faust, 1991)

1 Introduction

Campbell and Perron [9] have recently summarized much of the research on classical unit root econometrics. Their rules or insights are partly guidelines on how to proceed and partly warning signs on what to avoid. This paper formulates similar insights for the applied macroeconomist from a Bayesian perspective for two reasons. Firstly, Classical and Bayesian inference on unit roots can differ substantially. Secondly, a Bayesian perspective delivers more 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. Luckily, my job is easier than the one undertaken by Campbell and Perron [9], not only because we can rely on their terminology and because the literature on Bayesian methods is smaller, but also because Bayesian inference is simpler. At issue is usually only the choice of a suitable prior.

The method used ought to depend on the question asked. In section 2, I argue that macroeconomists are typically interested in some particular persistence property or medium-term forecasting property and that therefore Bayesian methods are especially suitable. After a review of Bayesian methodology for time series analysis in section 3 I discuss the choice of priors in section 4. 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 found to be reasonable provided they are modified by concentrating the prior mass for the time trend coefficient towards zero as the largest root approaches unity from below. Turning to the persistence issue in section 5, we discuss that classicists often do not take the uncertainty about the underlying parameter into account in a satisfactory manner, whereas Bayesian methods naturally do. I discuss that the tails of the predictive density can be sensitive to the prior treatment of explosive roots, an issue on which classical methods are typically mute.

In summary, a Bayesian approach is easy and practical to employ and often delivers more natural answers to macroeconomic questions.

2 Questions

Most macroeconomic research is aimed at answering some particular question. For example a researcher may begin by asking

1. How big will GNP be in the long-run?
2. What is the effect of monetary policy on output?
3. Does the expectations theory of interest rates hold?
4. Is there excess volatility on the stock market?
5. Is there convergence in per capita output across countries?

The researcher then progresses to answer the question using the methods available to the modern macroeconomist. Obviously the question ought to determine the methods used.

Insight 1 *Formulate your question.*

Econometric tools and formal inference are often helpful for answering many questions arising in macroeconomics. Broadly speaking, there are two methodologies: classical and Bayesian[1] (for a deeper discussion of methodological issues, see Hendry and Richard [36], Hendry [37], Leamer [57], Pagan [74], Phillips [78], Sims [98], Berger and Wolpert [5], Poirier [83, 84], Rust [87], Pagan [75], Geweke [30], Zellner [115], Pagan and Wickens [76] and Kydland and Prescott [53]). Classical methods view the true parameter of interest as unknown but fixed and its estimator as random. Bayesian methods by contrast take the data and estimators computed thereof as given and regard the true parameter as random. Even though these two methods often result in similar calculations and conclusions otherwise, they differ sharply on possibly nonstationary time series[2] as was pointed out by Sims [100].

[1] Most researchers use classical methods probably because that method was the first if not the only one they were taught. Nonetheless they may remember how hard it was at first to translate rejecting the null hypothesis at a 5 % significance level into thinking "If the null hypothesis is true, we will do that at most 5 % of the time" rather than "The null hypothesis is true with a probability of 5 %". It therefore seems that most people are born as Bayesians and converted to Classicists by their first class in statistics.

[2] Recently, Phillips and Ploberger [81, 82] and Elliott and Stock [25] have attempted to reconcile classical and Bayesian inference in time series models.

Hence, in order to determine the most suitable method, the question asked needs to be made operational. The list of questions above, for example, is reduced to questions like "Is GNP trend-stationary or difference-stationary?", "Are Money and Output cointegrated?", "Are interest rates of various horizons cointegrated?", "Is there a unit root in dividends?", or "Are output of country A and output of country B cointegrated?" or at least it is claimed that it is necessary to answer these questions first before progress can be made on the "true" question of interest. While Bayesian posterior odds ratios provide a suitable tool for performing these tests (see e.g De Jong and Whiteman [19, 20], Schotman and Van Dijk [89, 91, 92], Phillips [79], Lubrano [66], Koop and Steel [49], Koop [50]), a classical econometrician typically reduces the questions even further to "Can we reject the null hypothesis that ... at the significance level of ..." and classical methods win the contest hands down by assumption.

These latter questions are only auxiliary, however. A classical econometrician needs answers to them in order to investigate the questions on the first list[3]. Unfortunately any test of the composite null of difference stationarity against trend stationarity or vice versa has power no greater than size in finite samples (see Campbell and Perron [9, Rule 8 and Rule 10] or Faust [27]), making the decision for either alternative insensible on the basis of such a test. This leads classical econometricians to conclude that

> *"for practical purposes it does not really matter if we label a difference-stationary process with coefficient $h_{\Delta y}(0)$ close to zero as a trend-stationary process, or if we label a trend-stationary process with extremely persistent shocks as a difference-stationary process. Indeed these kinds of errors may even have practical advantages."* (Campbell and Perron [9], p. 160).

In other words, Campbell and Perron [9] argue that it is of practical advantage to reject some particular null hypotheses practically always even though it is true.

Insight 2 *If it is of practical advantage for a classical econometrician to reject some null hypothesis in the majority of cases even though that hypothesis*

[3]This at least seems to be the main rationale for unit root tests put forward by their proponents. It turns out not to be true in a number of circumstances, see e.g Cochrane [14, 15] and Sims, Stock and Watson [102]. We will return to this issue again in section 5.

is true, then that null hypothesis is not the hypothesis of interest. The null hypothesis of trend stationarity and the null hypothesis of difference stationarity are often an example for this.

Indeed the issue which interests macroeconomists is typically some persistence property or medium-term forecasting property. An operational way to phrase the first list of questions should be

1. How big will GNP be in ten years?

2. How sizeable is the effect of a shock in monetary policy on GNP over the five years following that shock?

3. How much money can the goverment save in 30 years by replacing one 30 year T-Bond by 30 successive 1-year T-Bills[4]?

4. How much money can an investor make in ten years by investing 10000 $ on the stock market, following a contrarian strategy?

5. How large will the relative difference in per capita GNP between Japan and the US be in thirty years, say, compared to the relative difference today?

Roots near unity are important only insofar they sometimes imply sizable persistence of shocks. This point is not new and has been argued before[5] by e.g. Christiano and Eichenbaum [13], Cochrane [14, 15, 16] and Durlauf [24].

Insight 3 *For most questions, the issue is not one of trend-stationarity versus difference-stationarity or one of cointegration. For most questions the issue is some particular persistence property or some medium-term forecasting property of the data.*

[4] The Lucas critique (Lucas [67]) becomes obviously an important issue in attempting to answer questions phrased in this way. See also Lucas and Stokey [68].

[5] 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 [111], Campbell and Mankiw [8], Quah [86]).

Uncertainty about some medium-term forecast arises from three sources: model uncertainty, which is in general impossible to quantify[6], coefficient uncertainty and uncertainty about future innovations. While both methods, classical and Bayesian, recognize the uncertainty about future innovations, the Bayesian approach takes care of the coefficient uncertainty in a more natural and appealing way. Bayesian methods therefore seem more suitable to answer macroeconomic questions. We return to this issue in section 5 after a review of of Bayesian inference in time series models.

3 Methods

For a more thorough introduction, see e.g. Box and Tiao [3], Zellner [113], Leamer [54], Judge et al. [41], Berger [4], Broemeling [6], Berger and Wolpert [5], and West and Harrison [112]. Consider the m-variate model

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

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 $\epsilon_t, t = 1, \ldots, T$, size $m \times 1$, are independently and normally distributed[7] according to

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

A time trend should to be included to encompass both trend stationary and difference stationary possibilities as special cases:

[6]It may appear to some that this shouldn't be a problem since one could always test for some extra lag or for that nonlinear term not included originally, at least asymptotically. Unfortunately, data sets end and so does any such investigation. A Bayesian solution would be to impose a prior across all possible linear and nonlinear specifications of the model. Aside from mathematical problems arising from the infinite dimensionality of the appropriate "parameter" space, agreement on some appropriate prior on this space is unlikely.

[7]Bayesian Analysis is not restricted to normally distributed errors, of course. For other types of distributions or non-conjugate priors, numerical integration methods as in e.g. Kloek and Van Dijk [46], Naylor and Smith [72] and Geweke [29, 33, 34] are available. If the error distribution is unspecified, Bayesian limited information methods can be used, see e.g. Kwan [52], Kim [43].

Insight 4 *The model to be analyzed should contain all data generating processes of main interest as special cases.*

This rule mirrors rules 1 through 6 in Campbell and Perron [9]. The rule is not a blank check to include as many lags and nonlinear terms as possible but simply as a rule not to exclude the most interesting alternatives. Formulating the model as an autoregression with some maximal lag length as above (rather than a moving average with some maximal lag length, say) puts certain restrictions on the spectral density of Δy_t which matter for inferences about unit roots: this issue has been discussed at length elsewhere (see the survey in Campbell and Perron [9]). An autoregressive representation with finite lag length is often preferable since it implies rather smooth impulse response functions, a feature which is attractive a priori[8].

What is known at inference time is the data. What is unknown are the regression coefficients, summarized as $B = \left[B_{(1)}\ B_{(2)}\ \ldots\ B_{(k)}\ C_{(0)}\ C_{(1)}\right]'$, and the covariance matrix Σ or, alternatively, the precision matrix $H = \Sigma^{-1}$. A Bayesian first chooses a prior probability density function $\pi(B, H)$ in B and H. Multiplying this prior with the likelihood function, one obtains the posterior

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

with the constant of proportionality so that $\pi_{T,H}$ integrates to unity. The posterior is the key tool and used to answer the particular question at hand, see section 5.

How to choose a prior is discussed below in section 4. A few observations can already be made, however. Think about calculating $\pi_T(B,H)$ in two steps: first, find the "prior" $\pi_0(B,H)$ given the initial observations,

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

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

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

For the model (1), (2), the conditional likelihood function which appears in (5) turns out to have a particularly simple form.

[8]A more general investigation of this issue from a Bayesian perspective should be interesting.

Insight 5 *Given the data Y_t, $t = -k+1, \dots, T$, the conditional likelihood function as a function in B and H is proportional to a Normal-Wishart*[9] *density function. This is true regardless of whether there are unit roots, cointegrating vectors or explosive roots.*

This insight is the central message in Sims and Uhlig [104]: while 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). This is easy to see for a univariate AR(1) process and not much harder in general, see appendix B. Crudely speaking, the nonstationarity is in the data, not in the parameters, and the data is given at inference time. Conditioning on the covariance matrix Σ as well, the shape of the likelihood function is simply proportional to a normal distribution:

Insight 6 *Conventional t and F statistics and their conventional p-values, which are meaningful in summarizing the shape of the likelihood function under stationarity assumptions, are equally meaningful in summarizing the shape of the likelihood function when unit roots or even explosive roots are present.*

Since the likelihood function is a tool shared by both, classical as well as Bayesian econometricians, it is reasonable and meaningful to report standard t-statistics and standard p-values regardless of which kind of econometrics one adheres to. Just knowing the likelihood function alone already allows one to compute how much more likely it is that the data was drawn with parameters $(\tilde{B}, \tilde{H})$ rather than $(\breve{B}, \breve{H})$, if both are given even prior odds: simply compute the likelihood ratio at these two parameter points.

4 Priors

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

[9] See appendix A for a definition of a Normal-Wishart Distribution.

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", the prior expresses subjective prior beliefs about the parameters of interest.

Remarkably enough, in linear models similar to (32) 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 $\mid H \mid^{-(m+1)/2}$ (see e.g. Zellner [113, section 8.1]), or, more generally, a Normal-Wishart prior[10] in B and H. The posterior will then be Normal-Wishart as well. This result is still true in the model (1), (2), since Bayesian inference is conditional on the observed data (see appendix B for the precise formulas).

Insight 7 *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 insight explains the popularity of the Normal-Wishart prior. Two special cases are especially common (see appendix B for details and references). The first is the "flat prior" stated above, see e.g. DeJong [17]. The other is a prior which conditions on Σ, is otherwise normal in the coefficient matrix B around the random walk specification as mean, and softly excludes longer lags by tightening the prior around a value of zero for their coefficients, see e.g. the RATS Manual [23].

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 model. This was pointed out by Phillips [79] and discussed by Koop and Steel [48], Leamer [58], Kim and Maddala [42], Poirier [85], Schotman and van Dijk [91], Stock [106], DeJong and Whiteman [19], Sims [105], Phillips [80] and Schotman [93]. Parameter regions, where the data will easily distinguish nearby values are packed "denser" than others. It is therefore sensible to reparameterize the model so that the parameter space becomes "evenly packed" as measured by the data to be observed. In other words, reparameterize the model so that the Fisher information matrix function I is constant on the new parameter space and impose a flat prior then. Using calculus, this simply amounts to a prior

[10] See appendix A for a definition of a Normal-Wishart Distribution.

proportional to $| I |^{1/2}$, the square root of the determinant of Fisher's information matrix in the original parameterization. Such a prior assigns a lot of weight to parameter regions about which we expect the data to be very informative. This uninformative prior is called Jeffreys' prior, see Jeffreys [39, 40], Box and Tiao [3], Zellner [113], Leamer [54], several of the authors cited above and section 4.1 below. Jeffreys' prior has the additional and sometimes desirable property that it is immune to reparameterizations. While few endorse the uncritical use of Jeffreys' prior, it provides at least a helpful benchmark[11].

Insight 8 *Calculating Jeffreys' prior is helpful in choosing a suitable prior and understanding its implications.*

4.1 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 [79] and his discussants[12]: this debate is summarized below and extended with some recent results from Uhlig [110]. The aim here are the reasonableness and potential modifications of the popular Normal-Wishart prior.

For some univariate versions of (1), (2), Thornber [107], Zellner [113, section 7.1] and Phillips [79] calculate Jeffreys' priors. Consider the simple AR(1) model

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

Conditioning on the initial observation y_0, Phillips finds Jeffreys' prior or the "Critics' prior" (in accordance with Phillips [80]) to be

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

[11]Another candidate for priors representing "knowing little" are the maximal data information prior (MDIP) distributions, see Zellner [113, 114, 117] and Zellner and Min [119]. They are given by $\pi(\theta) \propto \exp(\iota(\theta))$, where $\iota(\theta) = \int p(y \mid \theta) \log(p(y \mid \theta))\, d\theta$ is Shannon's [95] measure of the information in the data distribution and where $p(y \mid \theta)$ is the probability density for the data y given the parameter vector θ. The MDIP is not invariant to reparameterisations.

[12]For the beginning debate about appropriate priors for multivariate models, see e.g. Kleibergen and van Dijk [45] and DeJong and Whiteman [21].

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 to the Critics'

– Insert figure 1 approximately here –

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.

Insight 9 *Conditional on non-explosive 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 big, once explosive roots are taken seriously. Figure 2 shows the same priors as

– Insert figure 2 approximately here –

figure 1 but restricted to $\rho \in [0.6; 1.1]$ rather than $\rho \in [0.5; 1]$. 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 $\rho > 1$.

Insight 10 *The Critics prior exhibits an explosive behaviour for explosive roots, i.e. roots above one in absolute value. To impose a rather flat or a declining prior for roots above one or a Normal-Wishart prior centered at the unit root or to set the prior to zero 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 [48], Leamer [58], Kim and Maddala [42], Poirier [85], Schotman and van Dijk [91], DeJong and Whiteman [19], and Sims [105], but also Phillips [80]. For example, DeJong and Whiteman [18, 19, 20] and DeJong [17] use a flat prior set to zero outside the interval $\rho \in [0.55\,,\,1.05]$.

Insight 11 *The prior belief of excluding too explosive roots 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 calculations were based on the conditional likelihood function rather than the exact likelihood function. Using the latter, Thornber [107] and Zellner [113] as extended by Uhlig [110] find Jeffreys' prior to be

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

The prior is written as a π_0-prior, see equation (4) for notation. This prior has an integrable singularity of order $(1-\rho)^{-1/2}$ at $\rho = 1$, see figures 1 and 2. Starting from this prior, Thornber [107], Zellner[13] [113, 114] and Lubrano [66] propose to use a Beta-Distribution on some interval $\rho \in [-\bar{\rho}, \bar{\rho}]$ as prior π: Thornber and Zellner suggest $\bar{\rho} = 1$ and Lubrano $\bar{\rho} > 1$. Berger in a seminar at Yale (1992) suggests extending the Jeffreys-Thornber-Zellner-prior "symmetrically" via the transformation $\tilde{\rho} = 1/\rho$ to the explosive region, resulting in the prior $\pi(\rho) = a(1-\rho^2)^{-1/2}$ for $\mid\rho\mid < 1$ and $\pi(\rho) = a(\rho^2 - 1)^{-1/2} \mid\rho\mid^{-1}$ for $\mid\rho\mid > 1$, where a is the appropriate integrating constant.

Sampling frequency may be an issue, see Sims [105]. Hartigan in a comment at Yale (1992) therefore suggests to use a sampling-frequency invariant prior $\pi_0(\rho) \propto 1/\left(\sigma\rho(-\log(\rho))^{\alpha}\right)$ for $0 \leq \rho \leq 1$ where α is a free parameter. Given σ, this prior has a non-integrable singularity at[14] $\rho = 1$ for $\alpha \geq 1$, whereas it has a non-integrable singularity at $\rho = 0$ for $\alpha \leq 1$. Since 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 [94]) computed Jeffreys' prior for a continuous

[13] Zellner investigates (6) including a constant: his calculations are easily modified for (6), however

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

time process $dy = \kappa y dt + \sigma dW$, $0 \leq t \leq T$, $y_0 = 0$ 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 $0 \leq \rho \leq 1$ (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 $\rho \in [0; 1]$, say.

The Maximal Data Information Prior or MDIP (see Zellner [113, 114, 117] and Zellner and Min [119]) for (6), using the exact likelihood function, is given by $\pi(\rho, \sigma) \propto (1-\rho^2)^{1/2}/\sigma$ and converges to zero as $\rho \to 1$. If the average Shannon information per observation of the entire sample is used to calculate the MDIP, one obtains $\pi(\rho, \sigma) \propto (1-\rho^2)^{1/(2T+2)}/\sigma$, which converges to a flat prior as the sample size T grows to infinity.

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

Insight 12 *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 substantially impact answering some questions (see e.g. the Bayesian unit root tests performed by computing the posterior probability $P(\rho \geq 1)$ in DeJong and Whiteman [18, 19, 20] and Phillips [79] or the predictive densities in section 5.). If feasible one should therefore follow Leamers [55, 56] suggestion, echoed in Poirier [83]:

Insight 13 *Analyze the sensitivity of the conclusions to a possible singularity at the unit root according to some of the priors discussed above. Likewise analyze the sensitivity of the conclusions to the prior treatment of explosive roots, if these roots are to be taken seriously.*

This is an opportunity rather than a pitfall of Bayesian inference. For example, Bayesian methods are well-suited to taking explosive roots seriously and exploring the fragility of answers to macroeconomic questions due to such roots. Classical methods by contrast usually consider at most roots above one which converge to one as the sample size increases to infinity[15].

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

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

with the restriction that $\beta = 0$ when $\rho = 1$ in order to exclude a quadratic time trend when there is a unit root. Phillips [80, 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 in the form of an unobserved components model

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

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

as suggested by e.g. Sims, Stock and Watson [102], Schotman and Van Dijk [91], Andrews [1] and Sims [105]. In this parameterization, the nature of the deterministic and the stochastic trend are nicely separated, whereas they are interdependent[16] in (11): since

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

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

For the unobserved components parameterization, Schotman and Van Dijk [91] calculate Jeffreys' prior $\pi_{0,J}$ using the conditional likelihood and obtain

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

[15]See however Chow [12, chapter 6].

[16]See Schotman [93] for a further discussion of such issues.

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=[17] (15). For $y_0 = \gamma$, one obtains again Phillips' prior $\pi_{0,P}$.

Using the exact rather than the conditional likelihood function for the unobserved components specification (13), Uhlig [110] calculates Jeffreys' prior and finds the corresponding prior $\pi_{0,J,\infty}$ for the "reduced form" specification (11) via the transformation rule. Conditional on (μ,ρ,σ), this prior prescribes a particular normal distribution for β, the coefficient on the time trend:

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

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

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

conditional on (μ,ρ,σ). We suggest to 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

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

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

Calculating the MDIP (see Zellner [113, 114, 117] and Zellner and Min [119]) results in $\pi \propto (1-\rho^2)^{1/2}/\sigma$ (respectively in $\pi \propto (1-\rho^2)^{1/(2T+2)}/\sigma$, when the average Shannon information per observation of the entire sample is used) 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): the MDIP is not immune to parameter transformations.

[17]Strictly speaking, the transformation rule is only valid for $\rho \neq 1$. Since we are for most parts interested in the shape of priors rather than their exact value at $\rho = 1$, we ignore this as a technicality.

Insight 14 *If a univariate autoregression including a constant and a linear time trend is to be analyzed and if there is concern about at most one unit root or explosive root, 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)*[18].

Once a prior is chosen, Bayesian tests of the unit root null hypothesis can be performed. DeJong and Whiteman [18, 19, 20] perform the test for

$$H_0 : \rho \geq 1 \quad \text{versus} \quad H_1 : \rho < 1, \tag{20}$$

calculating posterior masses for these regions using Normal-Wishart priors. Phillips [79] critized these tests since they are sensitive to the prior treatment of explosive roots, see our discussion above. As Schotman and Van Dijk [91] pointed out, the appropriate test for a unit root[19] often is

$$H_0 : \rho = 1 \quad \text{versus} \quad H_1 : |\rho| < 1, \tag{22}$$

however. To perform these tests, Bayesian posterior odd 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 since that can result in nonsensical infinite integrals for certain improper priors. For details, see e.g. Zellner [113], Schotman and Van Dijk [91].

[18] To calculate the posterior, calculate the posterior first using just the Normal-Wishart prior with the formulas (37) through (40). 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 one.

[19] A continuous prior will assign probability zero to any nullset and therefore probability zero to the region characterised by the unit root $\rho = 1$ in (11). However, using a Beveridge-Nelson MA-representation instead,

$$y_t = \gamma + \delta t + \alpha W_t + A(L)\epsilon_t, \tag{21}$$

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 since it is characterised by $\alpha = 0$. The reason for this difference are restrictions on the spectral density for Δy_t introduced by the particular parameterization. It may be interesting to use Bayesian method directly for inference about the mass of the spectral density of Δy_t in some persistence-relevant region $[0, \nu]$, ν small.

5 Persistence and Forecasts

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

$$r(n) = \rho^n. \tag{23}$$

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

$$\hat{y}_{T+n} \equiv E\left[y_{T+n} \mid \rho, Y_T\right] = \rho^n y_T \tag{24}$$

and

$$Var\left[y_{T+n} \mid \rho, Y_T\right] = \begin{cases} \sigma^2 \dfrac{1-\rho^{2n}}{1-\rho^2}, & \text{if } \rho \neq 1, \\ n\sigma^2, & \text{if } \rho = 1. \end{cases} \tag{25}$$

The Bayesian and the Classicist differ however in their treatment of the uncertainty in ρ. We concentrate in our discussion on the forecasting exercise, but a persistence analsis via the impulse response functions clearly follows a similar logic (compare (23) to (24)).

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

$$\hat{\rho} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} \tag{26}$$

and

$$\sigma_\rho^2 = \frac{\sigma^2}{\sum_{t=1}^T y_{t-1}^2}. \tag{27}$$

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

$$\Delta y_t = \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0,1)\ i.i.d.,\ t = 1, \ldots, T. \tag{28}$$

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

$$E\left[y_{T+n} \mid Y_T\right] = y_T \tag{29}$$

and

$$Var\left[y_{T+n} \mid Y_T\right] = n \tag{30}$$

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

– Insert figure 3 approximately here –

ends up being too sure about his forecast: even though he does not actually know whether there is a unit root or not, the forecast distribution assumes that he does and does 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[20]. Alternatively he could evaluate the distribution of the n-step ahead forecast, taking into account the pretesting procedure. However this is rarely done (see e.g. Campbell and Perron [9], Cochrane [14] and also Fair [26] and Sampson [88]) and almost never for forecasts many periods into the future, despite the fact that the informal rational for pretesting for unit roots is persistence. Even if such an analysis was done more often, the resulting forecast distribution of the sophisticated classicist would depend on the unknown parameter, which is not known when performing the forecast, but would not depend on the data, which is known at forecasting time. It seems more natural to state the uncertainty about a forecast by including parameter uncertainty, but 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

[20]In particular, Andrews [1] has recently argued to correct 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 $\tilde{r}(n) = \tilde{\rho}^n$.

posterior[21] is given by

$$\rho \sim \mathcal{N}\left(\hat{\rho}, \sigma_\rho^2\right) \tag{31}$$

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 distributions for $\hat{y}_{T+n}$ or, alternatively, the uncertainty about the $r(n)$, the value of the impulse response function at step n, can be calculated directly[22] and are plotted in figure 4.

– Insert figure 4 approximately here –

Note that the distributions becomes 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. Adding the uncertainty about future ϵ_t, one obtains the predictive densities[23] for $y_{T+n} \mid Y_T$. They are plotted in figure 5 for $n = 1, \ldots, 10$ steps

– Insert figure 5 approximately here –

ahead. They are not normal, skewed and quite different[24] from the distributions in figure 3. Clearly the mass assigned by the posterior to the region

---

[21] If the prior is fixed and continuous or at least converges to a continuous proper prior in the limit as the sample sise increases to infinity, the posterior will in fact be asymptotically normal, as Kim [44] has shown. Thus a normal-shaped posterior is a good benchmark from which to analyse the forecasting issue.

[22] 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 arising from negative ρ. The distribution for $r(n)$, the value of the impulse response function at step n, is given by substituting $y_T = 1$ in this expression.

[23] For general results regarding these predictive densities, see Chow [11], who has explicitely calculated the mean for multivariate autoregressive model predictions if a Normal-Wishart prior is used, and Koop, Osiewalski and Steel [51].

[24] Given any posterior, these distributions converge for $n \to \infty$ to a limit which is given by the "weighted sum" of all stationary distributions $F_\rho(y)$ for y drawn from (6) with $\mid \rho \mid < 1$, 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 $\mid \rho \mid \geq 1$. 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 [1] analysis.

$\rho \geq 1$ matters for evaluating the uncertainty about the n-step ahead forecasts.

Predictive densities using any of the other priors discussed in the previous section can be created similarly: figure 6 and figure 7 shows five of them and

– Insert figure 6 approximately here –

– Insert figure 7 approximately here –

the resulting posteriors together for comparison. Figure 8 and 9 shows the forecast distribution for y_t 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 pos-

– Insert figure 8 approximately here –

terior) to $\rho \in [0; 2.0]$, whereas we restricted the prior to $\rho \in [0; 1.1]$ in figure

– Insert figure 9 approximately here –

9. In particular the tails (and thus for example the posterior probabilities for specific events like a downturn in GNP, see Zellner, Hong and Gulati [116]) depend on this restriction. In figure 10 the prior $\pi_{0,J,\infty}$ given in equation (9)

– Insert figure 10 approximately here –

is used to calculate the distributions 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 which persists into the posterior. Figure 11 is similar to figure 5 and shows the forecast distribution for y_{T+n},

– Insert figure 11 approximately here –

when the prior $\pi_{0,J,\infty}$ is used and restricted to $[0; 2]$. Figure 12 shows the

– Insert figure 12 approximately here –

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 forecast distributions rather than impulse response functions as in figure 10, but the particular exclusion of explosive roots does.

Insight 15 *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.*

There is more uncertainty about the forecast remaining in any of these figures than in figure 3, since the Bayesian takes into account the uncertainty about the underlying coefficient whereas a classicist often does not when using unit root tests as a pretesting device.

Insight 16 *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.*

It may therefore be advisable for calculating predictive densitities in practice to start with a reasonably general model and to stick with it (for a Bayesian critic of this position, see e.g. Zellner [118]. Sometimes, of course, the goal is a parsimonious, "stylized" description of the data itself). Do not pretest for unit roots or cointegrating relationships. Do not pretest for lag length either, but rather downweigh the importance of lagged variables by imposing tight priors around a value of zero for the coefficients on these variables (see appendix B for references).

6 Extensions

Much remains to be done. A Bayesian analysis of persistence measures as surveyed by e.g. Durlauf [24, pp. 72-77] is desirable. Some have claimed that the broad features of major macroeconomic time series can be explained by breaks in trends rather than unit roots (see e.g. Perron [77]): the Bayesian analysis of this issue should be extended (see e.g. Broemeling and Tsurumi [7] and Zivot and Phillips [120]). ARCHness is considered to be important by many and may interact with possible nonstationarities (see e.g. Geweke [31, 32] and Uhlig [109] for a Bayesian treatment of ARCHness). It may be that fractionally integrated processes are more suitable for evaluating persistence properties, see e.g. Lo [65]: a Bayesian analysis of this issue is desirable.

7 Conclusions

In this paper, we summarized the recent literature on the Bayesian treatment of unit roots into practical insights for applied macroeconomists. We argued that macroeconomists are typically interested in some persistence property or medium-run forecasting property and that therefore Bayesian methods are especially suitable.

We reviewed Bayesian methodology for time series analysis in section 3 and discussed the choice of priors in section 4. We summarized and extended the recent discussion about appropriate choices of priors. Recognizing a consensus distaste for explosive roots, we find the popular Normal-Wishart priors centered at the unit root reasonable provided they are modified by concentrating the prior mass for the time trend coefficient towards zero as the largest root approaches unity from below.

As for persistence and medium-term forecasting, we discussed that classical methods often do not take the uncertainty about the underlying parameter into account in a satisfactory manner, whereas Bayesian methods naturally do. We discussed that the tails of the predictive densities can be sensitive to the prior treatment of explosive roots and noted that classical methods are typically mute on this issue.

In summary, a Bayesian approach is easy and practical to employ using the insights in this paper. Such an approach often delivers more natural answers to questions that macroeconomists are interested in.

Appendix

A Some distributions and their properties.

More information about the distributions described below can be found e.g. in Muirhead [71] or Zellner [113]. A Wishart distribution $\mathcal{W}_m(\Omega, \nu)$, $\nu \geq m$ specifies the following density for the $m(m+1)/2$ distinct elements of a positive definite $m \times m$ random matrix X:

$$f_W(X \mid \Omega, \nu) = \frac{\mid X \mid^{(\nu-m-1)/2}}{2^{m\nu/2}\Gamma_m\left(\frac{\nu}{2}\right) \mid \Omega \mid^{\nu/2}} \exp\left(-\frac{1}{2} tr\left(\Omega^{-1}X\right)\right),$$

where Γ_m is the multivariate gamma function, defined in Muirhead [71], Definition 2.1.10. If $X \sim \mathcal{W}_m(\Omega, \nu)$, then $E[X] = \nu\Omega$. Note that $\mathcal{W}_1(1, \nu) = \chi^2_\nu$, the chi-squared distribution with ν degrees of freedom.

Let $l \geq 1$ be an integer: in the main text we usually have $l = km + 2$. Let N be $l \times l$ and positive definite, let S be $m \times m$ and positive definite, let $\bar{B}$ be a $l \times m$-dimensional matrix and let $\nu \geq m$. The Normal-Wishart distribution $\phi_{NW}(B, H \mid \bar{B}, N, S, \nu)$ for a $l \times m$-dimensional matrix B and a $m \times m$ matrix H is given by the density

$$\begin{aligned} \phi_{NW}(B, H \mid & \bar{B}, N, S, \nu) = \\ & \kappa(N, S, \nu, m) \mid H \mid^{l/2} \\ & \exp\left(-\tfrac{1}{2} vec(B - \bar{B})' \left[H \otimes N\right] vec(B - \bar{B})\right) \\ & \mid H \mid^{(\nu-m-1)/2} \exp\left(-\tfrac{1}{2}\nu \; tr HS\right) \end{aligned}$$

where $\kappa(N, S, \nu, m)$ is the appropriate integrating constant. The Normal-Wishart distribution specifies, that H follows a Wishart distribution $\mathcal{W}_m(S^{-1}/\nu, \nu)$ with mean[25] S^{-1}, and that conditional on H, the matrix B in its vectorized form $vec(B)$ follows a Normal Distribution $\mathcal{N}\left(vec(\bar{B}), H^{-1} \otimes N^{-1}\right)$.

[25] In the main text, $H = \Sigma^{-1}$ is the precision matrix and e.g. $S = \hat{\Sigma}$. Thus, perhaps more suggestively $E[\Sigma^{-1}]^{-1} = \hat{\Sigma}$.

B Rewriting the Likelihood Function.

Write $B = \left[B_{(1)}\ B_{(2)}\ \ldots\ B_{(k)}\ C_{(0)}\ C_{(1)} \right]'$ and let $\beta = vec(B)$ be its columnwise vectorization. Let $X_t = \left[Y'_{t-1}\ Y'_{t-2}\ \ldots\ Y'_{t-k}\ 1\ t \right]'$ and stack the observations via $\mathbf{Y} = [\,Y_1\ \ldots\ Y_T\,]'$, $\mathbf{X} = [\,X_1\ \ldots\ X_T\,]'$ and $\epsilon = [\,\epsilon_1\ \ldots\ \epsilon_T\,]'$. We can then rewrite equation (1) as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \epsilon. \tag{32}$$

Conditional on the initial observations Y_t, $t = 1 - k, \ldots, 0$, the likelihood function in the coefficient matrix B (or coefficient vector β) and the precision matrix $H = \Sigma^{-1}$ can be written as

$$\begin{aligned} L(B, H \mid \mathbf{Y}) &= (2\pi)^{-mT/2} \mid H \mid^{T/2} \exp\left(-\frac{1}{2}(\beta - \hat{\beta})' \left[H \otimes \mathbf{X}'\mathbf{X} \right] (\beta - \hat{\beta}) \right) \\ & \quad \exp\left(-\frac{T}{2} tr\, HS \right), \end{aligned} \tag{33}$$

where

$$\hat{\beta} = vec(\hat{B}),\ \hat{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{34}$$

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

$$\hat{\Sigma} = \frac{1}{T}(\mathbf{Y} - \mathbf{X}\hat{B})'(\mathbf{Y} - \mathbf{X}\hat{B}) \tag{35}$$

are the MLE's for β or B and Σ (see e.g. Zellner [113, section 8.1]). Thus, as a function in B and H, the likelihood function is proportional to a Normal-Wishart density[26] $\phi_{NW}(B, H \mid \hat{B}, N, (T/\nu)\hat{\Sigma}, \nu)$ with $\nu = T - (k-1)m - 1$. Given any prior $\pi_0(B, H)$ over the parameters B and H, conditional on the initial observations Y_t, $t = 1 - k, \ldots, 0$, equation (5) to calculate the posterior can therefore be written as

$$\pi_T(B, H) \propto \pi_0(B, H)\phi_{NW}(B, H \mid \hat{B}, N, (T/\nu)\hat{\Sigma}, \nu), \tag{36}$$

where the constant of proportionality is determined in such a way that π_T integrates out to unity. Sometimes it is desirable to describe the posterior as a function in B and Σ rather than in B and H. This is easy to do: replace H in

[26] See appendix A for a definition of a Normal-Wishart Distribution.

the formulas for the densities by Σ^{-1} and multiply the right hand side of (36) by $\mid \Sigma \mid^{-(m+1)}$, the Jacobian of the transformation $\Sigma^{-1} \to \Sigma$. One obtains an inverted Normal-Wishart rather than a Normal-Wishart distribution, see Zellner [113, section 8.1].

As pointed out in section 4, the posterior π_T is easy to calculate, if π_0 is a Normal-Wishart prior. The following proposition is a straightforward and well-known generalization of Leamer [54, Theorem 3.9]:

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

$$\phi_{NW}\left(B, H \mid B_0, N_0, S_0, \nu_0\right),$$

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

$$\phi_{NW}\left(B, H \mid B_T, N_T, S_T, \nu_T\right)$$

as well, where

$$\begin{aligned}
\nu_T &= T + \nu_0, && (37)\\
N_T &= N_0 + \mathbf{X}'\mathbf{X}, && (38)\\
\bar{B}_T &= N_T^{-1}\left(N_0 B_0 + \mathbf{X}'\mathbf{X}\hat{B}\right) && (39)\\
S_T &= \frac{\nu_0}{\nu_T} S_0 + \frac{T}{\nu_T}\hat{\Sigma} + && (40)\\
&\quad \frac{1}{\nu_T}\left(\hat{B} - \bar{B}_0\right)' N_0 N_T^{-1}\mathbf{X}'\mathbf{X}\left(\hat{B} - \bar{B}_0\right)
\end{aligned}$$

Proof: *Adapt the proof of Leamer [54, Theorem 3.9].* •

In particular, a "flat prior" is sometimes used (see e.g. DeJong [17], Broemeling [6], Broemeling and Tsurumi [7] and Koop [47]): $\bar{B}_0$ 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 $\mid H \mid^{-(m+1)/2}$ 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 $\Sigma = S_0 =$ [illegible] 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 [96, 97, 99, 101], Litterman [60, 61, 62, 63, 64], Doan, Litterman and Sims [22], Highfield [38], the RATS manual [23], Lütkepohl [69], Canova [10], and Uhlig [109]. This prior is sometimes referred to as the "Minnesota prior". The Software RATS [23] performs the required calculations.

One can understand these calculations (at least when conditioning on H) as signal extraction: 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 which are much more "imprecise" than the data about the important aspects of the parameters.

References

[1] Andrews, D. W. K., "Exactly Unbiased Estimation of First Order Autoregressive Unit Root Models," Cowles Foundation Discussion Papers 975 (1991)

[2] Amemyia, Takeshi, *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts (1985)

[3] Box, G. E. P. and G. C. Tiao, *Bayesian Inference in Statistical Analysis,* Addison Wesley, Reading, Massachusetts (1973)

[4] Berger, J.O., *Statistical Decision Theory and Bayesian Analysis,* 2nd. edition, Springer-Verlag, New York (1985)

[5] Berger, J.O. and Robert L. Wolpert, *The Likelihood Principle,* Institute of Mathematical Statistics, Lecture Notes Vol. 6, Hayward California (1988)

[6] Broemeling, Lyle D., *Bayesian Analysis of Linear Models,* Marcel Dekker, Inc., New York (1985)

[7] Broemeling, Lyle D., and Hiroki Tsurumi, *Econometrics and Structural Change,* Marcel Dekker, Inc., New York (1987)

[8] Campbell, John Y. and N. Gregory Mankiw, "Are Output Fluctuations Transitory?" Quarterly Journal of Economics, vol. 102 (1987), 857-880

[9] Campbell, John Y. and Pierre Perron, "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," NBER Macroeconomics Annual 1991, Olivier Jean Blanchard and Stanley Fisher, eds., The MIT Press, pp. 141-201

[10] Canova, Fabio, "Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting," draft, Department of Economics, European University Institute (1991)

[11] Chow, Gregory C., "Multiperiod Predictions from Stochastic Difference Equations by Bayesian Methods," Econometrica, vol. 41, no. 1 (1973), 109-118

[12] Chow, Gregory C., *Analysis and Control of Dynamic Economic Systems,* 2nd edition, Robert E. Krieger Publishing Company, Malabar, Florida (1986)

[13] Christiano, Lawrence J. and Martin Eichenbaum, "Unit Roots in GNP: Do We Know, And Do We Care?" Carnegie-Rochester Conference Series on Public Policy 32 (1990), 7-62

[14] Cochrane, John H., "How big is the random walk in GNP?" Journal of Political Economy, vol. 96 (1988), 893-920

[15] Cochrane, John H., "A critique of the application of unit root tests," Journal of Economic Dynamics and Control, vol 15 (1991), 275-284

[16] Cochrane, John H., "Comment," NBER Macroeconomics Annual 1991, Olivier Jean Blanchard and Stanley Fisher, eds., The MIT Press, pp. 141-201

[17] DeJong, D.N. "Co-Integration and Trend-Stationarity in Macroeconomic Time Series: Evidence from the Likelihood Function," draft (1991), Journal of Econometrics, forthcoming,

[18] DeJong, D.N. and C.H. Whiteman, "The Temporal Stability of Dividends and Stock Prices: Evidence from the Likelihood Function," American Economic Review, vol. 81, no. 3 (1991), 600-617,

[19] DeJong, D.N. and C.H. Whiteman, "The Case for Trend-Stationarity is Stronger Than We Thought," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 413-422,

[20] DeJong, D.N. and C.H. Whiteman, "Trends and Random Walks in Macroeconomic Time Series: A Reconsideration Based on the Likelihood Principle," Journal of Monetary Economics, forthcoming

[21] DeJong, D.N. and C.H. Whiteman, "Modelling Stock Prices Without Pretending to Know How to Induce Stationarity," draft, University of Iowa (1992)

[22] Doan, Thomas, Robert Litterman, and Christopher Sims, "Forecasting and conditional projection using realistic prior distributions," Econometric Reviews, Vol. 3, No. 1 (1984), p. 1-100,

[23] Doan, Thomas, *RATS Manual*, VAR Econometrics, 1800 Sherman Ave. Suite 612, Evanston, IL 60201 (1990)

[24] Durlauf, Steven N., "Output Persistence, Economic Structure, and the Choice of Stabilization Policy," Brookings Papers on Economic Activity 2 (1989), 69-136

[25] Elliot, Graham, and James H. Stock, "Inference in Time Series Regressions when there is Uncertainty about Whether a Regressor Contains a Unit Root," draft, Harvard University (1992)

[26] Fair, Ray C., "Estimating Event Probabilities from Macroeconometric Models Using Stochastic Simulation," NBER Technical Working Paper 111 (1991)

[27] Faust, Jon, "Near Observational Equivalence: Generic Concepts and Peculiar Implications for Unit Root Tests," draft, International Finance Division of the Federal Reserve Board (1991, rev. 1992)

[28] Florens, Jean-Pierre, Michel Mouchart and Sophie Larribeau-Nori, "Bayesian Encompassing test of a Unit Root Hypothesis," draft, University of Toulouse (1992)

[29] Geweke, J., "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference," Journal of Econometrics, vol. 38 (1988), 73-90

[30] Geweke, J., "Comment of Poirier: Operational Bayesian Methods in Econometrics," Journal of Economic Perspectives, vol. 2, no. 1 (1988), 159-166,

[31] Geweke, John, "Exact inference in models with autoregressive conditional heteroskedasticity," Chapter 5 in "Dynamic econometric modelling," W.A. Barnett, E.R. Berndt, H. White, eds., Cambridge University Press (1988), p. 73-105

[32] Geweke, John, "Exact predictive densities for linear models with ARCH disturbances," Journal of Econometrics, Vol. 40 (1989), p. 63-86

[33] Geweke, J., "Bayesian Inference in Econometric Models Using Monte Carlo Integration," Econometrica, vol. 57 (1989), 1317-1339

[34] Geweke, J., "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," Federal Reserve Bank of Minneapolis, Research Department Staff Report 148 (1991)

[35] Geweke, J., "Construction of Priors for Economic Time Series," draft, University of Minnesota (1992)

[36] Hendry, David F. and Jean-Francois Richard, "On the Formulation of Empirical Models in Dynamic Econometrics," Journal of Econometrics, vol. 20 (1982), 3-33

[37] Hendry, D.F., "Econometric Methodology: A Personal Perspective," in T. Bewley, ed., *Advances in Econometrics,*, vol. 2, Cambridge University Press, Cambridge (1987)

[38] Highfield, Richard A., "Forecasting with Bayesian State Space Models," Doctoral Dissertation, Graduate School of Business, University of Chicago (1986)

[39] Jeffreys, H., "An invariant form for the prior probability in estimation problems," Proceedings of the Royal Society of London, series A, vol. 186 (1946), 453-461

[40] Jeffreys, H., *Theory of Probability*, 3rd. ed., Oxford University Press, London (1961)

[41] Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lüetkepohl, Tsoung-Chao Lee, *The Theory and Practice of Econometrics*, 2nd edition, John Wiley and Sons, New York (1985)

[42] Kim, I.-M and G.S. Maddala, "Flat Priors vs. Ignorance Priors in the Analysis of the AR(1) Model," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 375-380,

[43] Kim, Jae-Young, "A Limited Information Bayesian Procedure: I. Theoretical Investigation," draft, University of Minnesota (1991)

[44] Kim, Jae-Young, "Large Sample Properties of Posterior Densities in a Unit-Root Time Series Model," draft, University of Minnesota (1991)

[45] Kleibergen, Frank and Herman K. van Dijk, "On the Shape of the Likelihood/Posterior in Cointegration Models," draft, Econometric Institute and Tinbergen Institute, Erasmus University, Rotterdam (1992)

[46] Kloek, T., and H.K. Van Dijk, "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," Econometrica, vol. 46 (1978), 1-20

[47] Koop, Gary," Aggregate shocks and macroeconomic fluctuations: a predictive approach," working paper, Boston University (1990)

[48] Koop, G. and 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, vol. 6, no. 4 (1991), 365-370,

[49] Koop, G. and 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),

[50] Koop, G., "Objective Bayesian Unit Root Tests," Journal of Applied Econometrics, vol. 7 (1992), 65-82

[51] Koop, G., Jecek Osiewalski and Mark F. J. Steel, "Bayesian Long-Run Prediction in Time-Series Models," draft, Boston University (1992)

[52] Kwan, Yum-Keung, "Bayesian Analysis with an Unknown Likelihood Function - A Limited Information Approach," draft, Graduate School of Business, University of Chicago (1990),

[53] Kydland, Finn E. and Edward C. Prescott, "The Econometrics of the General Equilibrium Approach to Business Cycles," Federal Reserve Bank of Minneapolis, Staff Report 130 (1990)

[54] Leamer, E. E., *Specification Searches,* John Wiley & Sons, New York (1978)

[55] Leamer, E. E., "Let's take the con out of econometrics," American Economic Review, vol 73 (1983), 31-43

[56] Leamer, E. E., "Sensitivity analysis would help," American Economic Review, vol. 75 (1983), 308-313

[57] Leamer, E. E., "Econometric Methaphors," in T. Bewley, ed., *Advances in Econometrics,*, vol. 2, Cambridge University Press, Cambridge (1987)

[58] Leamer, E.E., "Comment on 'To Criticize the Critics," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 371-374,

[59] Lindley, D.V., "The Use of Prior Probability Distributions in Statistical Inference and Decision," Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability 1 (1961), 453-468

[60] Litterman, Robert," Techniques of forecasting using vector autoregressions", Federal Reserve Bank of Minneapolis, working paper 115 (1979),

[61] Litterman, Robert,"A Bayesian procedure for forecasting with vector autoregressions," working paper, MIT (1980)

[62] Litterman, Robert,"Forecasting and policy analysis with Bayesian vector autoregression models," Federal Reserve Bank of Minneapolis, Quarterly Review (1984), p. 30-41

[63] Litterman, Robert,"Specifying vector autoregressions for macroeconomic forecasting," Federal Reserve Bank of Minneapolis, Staff Report 92 (1984)

[64] Litterman, Robert,"Forecasting with Bayesian vector autoregressions - five years of experience," Journal of Business and Economic Statistics, vol. 4 (1986), 25-38

[65] Lo, Andrew W., "Long-Term Memory in Stock Prices," Econometrica, vol. 59, no. 5 (1991), 1279-1314

[66] Lubrano, M., "Testing for Unit Roots and Cointegration in a Bayesian Framework," draft, CREQE, Marseille (1991)

[67] Lucas, Robert E., Jr., "Econometric Policy evaluation: A Critique," in *The Phillips Curve and Labor Markets,* K. Brunner and A. H. Meltzer,

eds, Carnegie-Rochester Conference Series no. 1, North Holland, New York (1976)

[68] Lucas, Robert E., Jr. and Nancy L. Stokey, "Optimal Fiscal and Monetary Policy in an Economy without Capital," Journal of Monetary Economics, vol. 12 (1983), 55-94

[69] Lütkepohl, Helmut, *Introduction to Multiple Time Series Analysis*, Springer Verlag, Berlin (1990)

[70] McCulloch, R.E. and R.S. Tsay, "Bayesian Inference of Trend- and Difference-Stationarity," draft, Graduate School of Business, University of Chicago (1992)

[71] Muirhead, Robb J., *Aspects of multivariate statistical theory*, John Wiley & Sons, Inc., New York (1982)

[72] Naylor, J. C. and A. F. M. Smith, "Econometric Illustrations of Novel Numerical Integration Strategies for Bayesian Inference," Journal of Econometrics, vol. 38 (1988), 103-125

[73] Nelson, Charles R. and Charles I. Plosser, "Trends and Random Walks in Macroeconomic Time Series," Journal of Monetary Economics, vol. 10 (1982), 129-162

[74] Pagan, Adrian, "Three Econometric Methodologies: A Critical Appraisal," Journal of Econometric Surveys, vol. 1, no. 1 (1987), pp. 3-23

[75] Pagan, Adrian., "Comment on Poirier: Dogma or Doubt?," Journal of Economic Perspectives, vol. 2, no. 1 (1988), 153-158,

[76] Pagan, A.R. and M.R. Wickens, "A Survey of Some Recent Econometric Methods," The Economic Journal, vol. 99 (1989), 962-1025

[77] Perron, Pierre, "The Great Crash, the Oil Shock, and the Unit Root Hypothesis," Econometrica, vol. 57, no. 6 (1989), 1361-1402

[78] Phillips, P. C. B., "Reflections on Econometric Methodology," draft, Cowles Foundation for Research in Economics, Yale University (1988)

[79] Phillips, P.C.B., "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 333-364,

[80] Phillips, P.C.B., "Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 3435-474,

[81] Phillips, Peter C.B. and Werner Ploberger, "Time Series Modeling with a Bayesian Frame of Reference: I. Concepts and Illustrations," draft, Cowles Foundation for Research in Economics, Yale University (1991)

[82] Phillips, Peter C.B. and Werner Ploberger, "Posterior Odds Testing for a Unit Root with Data Based Model Selection," draft, Cowles Foundation for Research in Economics, Yale University (1992)

[83] Poirier, D.J., "Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics," Journal of Economic Perspectives, vol. 2, no. 1 (1988), 121-144,

[84] Poirier, D.J., "The Subjectivist Response," Journal of Economic Perspectives, vol. 2, no. 1 (1988), 167-170,

[85] Poirier, D.J., "A Comment on 'To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends'," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 381-386,

[86] Quah, Danny, "The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds," Econometrica, vol. 60, no. 1 (1992), 107-118

[87] Rust, J., "Comment on Poirier: The Subjective Perspective of a 'Spiritual Bayesian'," Journal of Economic Perspectives, vol. 2, no. 1 (1988), 145-151,

[88] 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, vol. 6 (1991), 67-76

[89] Schotman, P.C. and H.K. van Dijk, "A Bayesian Analysis of the Unit Root in Real Exchange Rates," draft (1989), Journal of Econometrics, forthcoming

[90] Schotman, P.C. and H.K. van Dijk, "Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP," draft, Econometric Institute, Erasmus University (1990),

[91] Schotman, P.C. and H.K. van Dijk, "On Bayesian Routes to Unit Roots," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 387-402,

[92] Schotman, P.C., "When Unit Roots Matter: Excess Volatility and Excess Smoothness of Long Term Interest Rates," Institute for Empirical Macroeconomics, Discussion Paper 44, Federal Reserve Bank of Minneapolis (1991, rev. 1992)

[93] Schotman, Peter C., "Objective Inforamtion and Ignorance: NOtes on Some Priors in the Bayesian Analysis of the AR(1) Model," draft, Department of Finance, University of Limburg, Maastricht (1992)

[94] Schotman, Peter C., "Jeffreys' Prior in Continuous Time," draft, Princeton University (1992)

[95] Shannon, C. E., "The Mathematical Theory of Communication," Bell System Technical Journal (1948)

[96] Sims, C. A., "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," Journal of the American Statistical Association, vol. 67, no. 337 (1972), 169-175

[97] Sims, Christopher A.,"Are forecasting models usable for policy analysis?" Federal Reserve Bank of Minneapolis, Quarterly Review (1986), p. 2-15

[98] Sims, C. A. , "Making Economics Credible," in T. Bewley, ed., *Advances in Econometrics,*, vol. 2, Cambridge University Press, Cambridge (1987)

[99] Sims, Christopher A.,"Identifying policy effects," Federal Reserve Bank of Minneapolis, working paper 351 (1987),

[100] Sims, C.A., "Bayesian Skepticism on Unit Root Econometrics," Institute for Empirical Macroeconomics, Discussion Paper 3, Federal Reserve Bank of Minneapolis (1988) and Journal of Economic Dynamics and Control, vol. 12 (1988), 463-474

[101] Sims, Christopher A.,"A nine variable probabilistic macroeconomic forecasting model," Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, discussion paper 14 (1989),

[102] Sims, C.A., James H. Stock and Mark W. Watson, "Inference in Linear Time Series Models with Some Unit Roots," Econometrica, vol. 58, no. 1 (1990), 113-144

[103] Sims, C.A., "Modelling Trends," Discussion Paper ..., Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis (1990)

[104] Sims, C.A. and H. Uhlig, "Understanding Unit Rooters: A Helicopter Tour," Econometrica, vol. 59, no. 6 (1991), 1591-1600

[105] Sims, C.A., "Comment by Christopher A. Sims on 'To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends', by Peter C.B. Phillips," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 423-434,

[106] Stock, J.H.., "Bayesian Approaches to the 'Unit Root' Problem: A Comment," Journal of Applied Econometrics, vol. 6, no. 4 (1991), 413-422,

[107] Thornber, H., "Finite Sample Monte Carlo Studies: An Autoregressive Illustration," Journal of the American Statistical Association, vol. 62 (1967), 801-818

[108] Tsurumi, H. and H. Wago, "Bayesian Inference of Unit Root under an Unknown Order of an ARMA(p,q) Error Process and an Application to the Yen Foreign Exchange Rate," draft, Rutgers University (1992)

[109] Uhlig, H., "BVARTEC," draft, Princeton University (1991)

[110] Uhlig, H., "A Note on Jeffreys' Prior When Using the Exact Likelihood Function," draft, Princeton University (1992)

[111] Watson, Mark W., "Univariate Detrending Methods with Stochastic Trends," Journal of Monetary Economics, vol. 18 (1986), 49-75

[112] West, Mike and Jeff Harrison, *Bayesian Forecasting and Dynamic Models,* Springer-Verlag, New York (1989)

[113] Zellner, Arnold *An Introduction to Bayesian Inference in Econometrics,* John Wiley & Sons, New York (1971)

[114] Zellner, Arnold, "Maximal Data Information Prior Distributions", in *New Developments in the Applications of Bayesian Methods,* A. Aykac anc C. Brumat, eds., Amsterdam, North Holland Publ. Co. (1977), 211-32, also in *Basic Issues in Econometrics,* A. Zellner ed., University of Chicago Press (1984), 201-215

[115] Zellner, Arnold, "Bayesian Analysis in Econometrics," Journal of Econometrics, vol 37 (1988), 27-50

[116] Zellner, Arnold, Chansik Hong and Gaurang Mitu Gulati, "Turning Points in Economic Time Series, Loss Structures and Bayesian Forecasting," in *Bayesian and Likelihood Methods in Statistics and Econometrics*, S. Geisser, J.S. Hodges, S.J. Press and A. Zellner, eds., North-Holland, Amsterdam (1990)

[117] Zellner, Arnold, "Bayesian Methods and Entropy in Economics and Econometrics," in W.T. Grandy, Jr. and L.H. Schick (eds.), *Maximum Entropy and Bayesian Methods,* Kluwer Academic Publishers (1991), 17-31

[118] Zellner, Arnold, "Time Series Analysis, Forecasting and Econometric Modeling: The Structural Econometric Modeling, Time Series Analysis (SEMTSA) Approach," draft, Graduate School of Business, University of Chicago (1991)

[119] Zellner, Arnold and Chung-ki Min, "Bayesian Analysis, Model Selection and Prediction," draft, Graduate School of Business, University of Chicago (1992)

[120] Zivot, Eric and Peter C. B. Phillips, "A Bayesian Analysis of Trend Determination in Economic Time Series," Cowles Foundation Discussion Paper 1002 (1991)

[121] Zivot, Eric, "A Bayesian Analysis of the Unit Root Hypothesis Within an Unobserved Components Model," draft, Wellesley (1992)

Figure 1

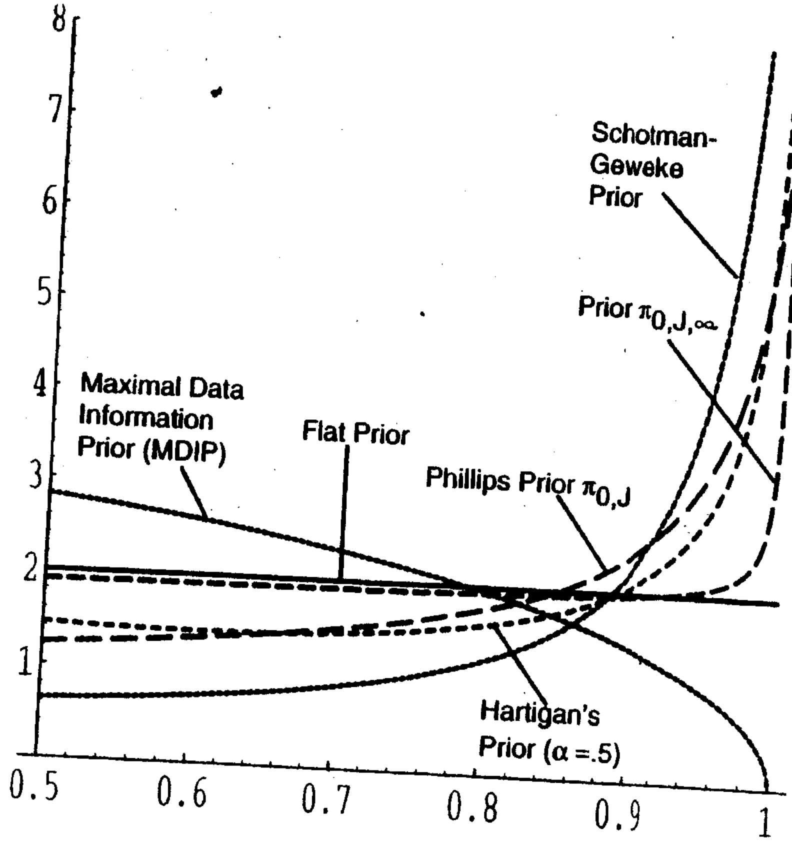


Figure 2

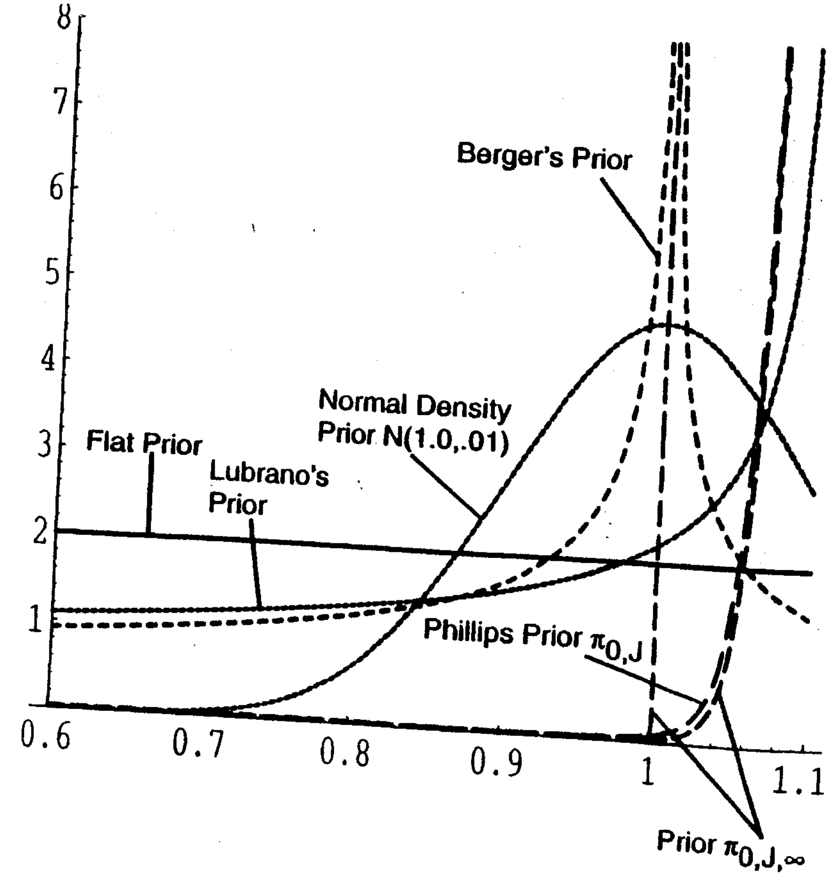


Figure 3

(Classicists' n-step ahead distribution of y_t)

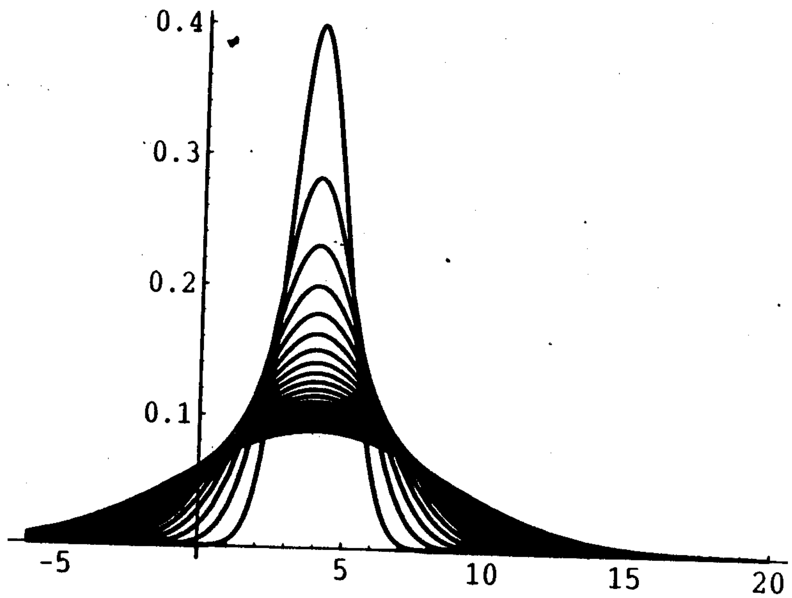


Figure 4

(Bayesian n-step ahead distribution of $E[y_t|\rho,\text{data}]$)

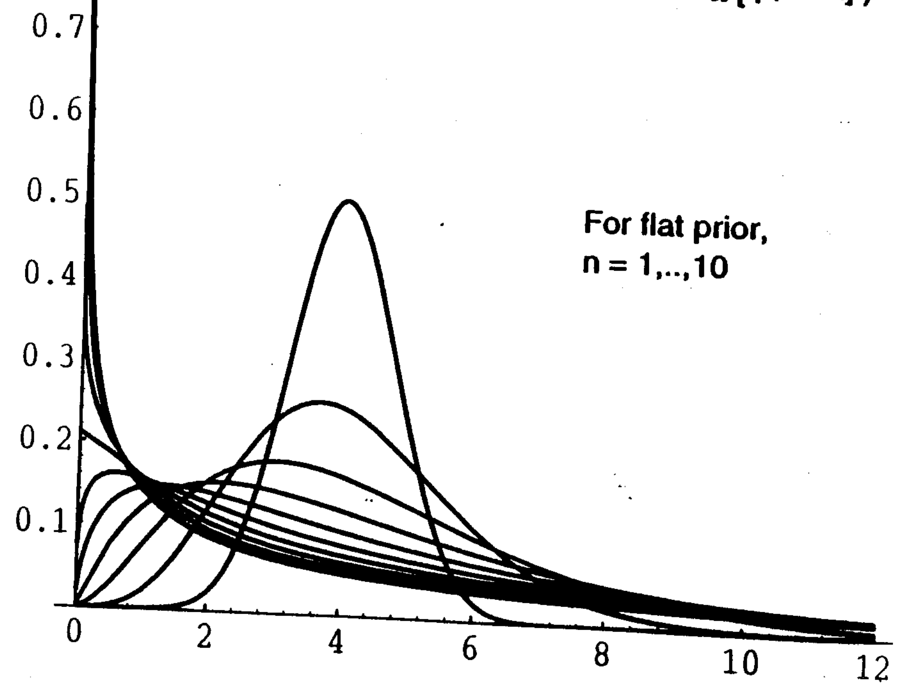


Figure 5

(Predictive densities for y_t, n steps ahead)

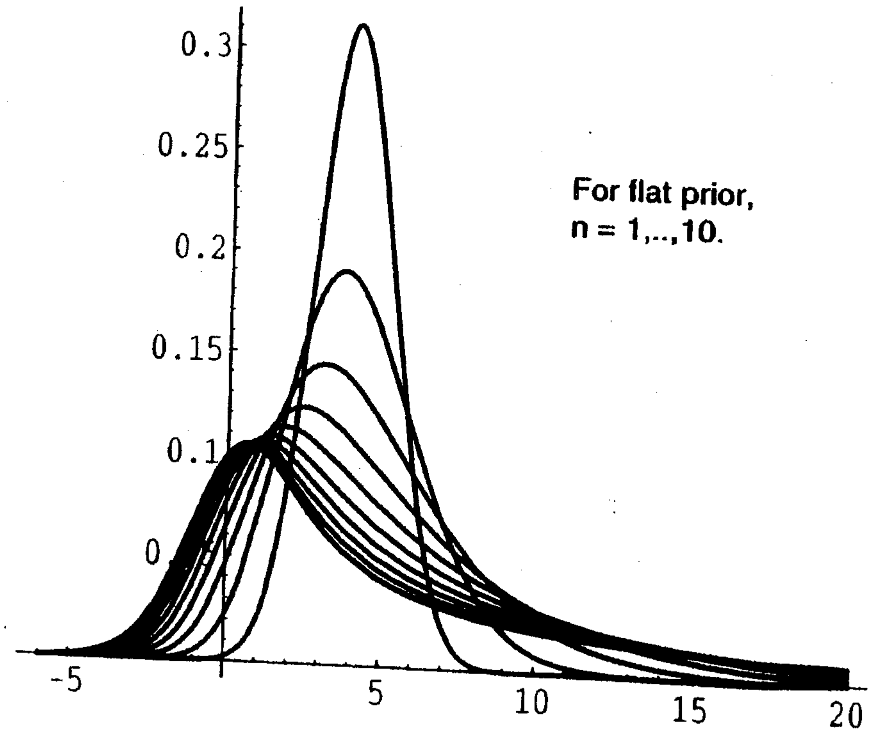


Figure 6
(Priors)

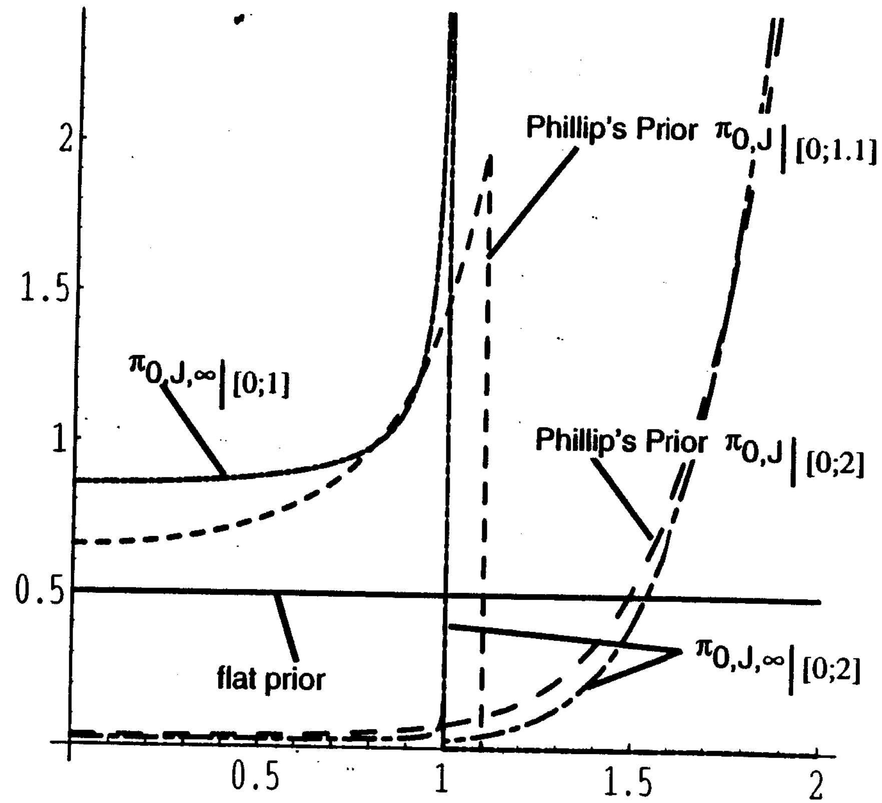


Figure 7
(Posteriors to priors acc. to label)

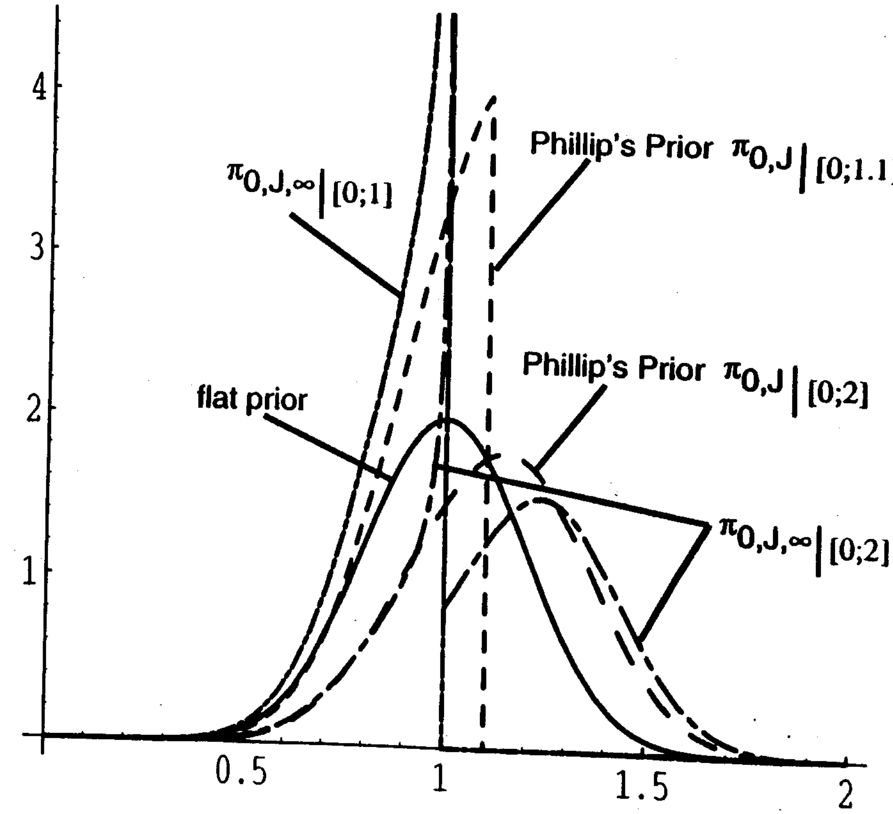


Figure 8

(Predictive densities for y_t, n steps ahead)

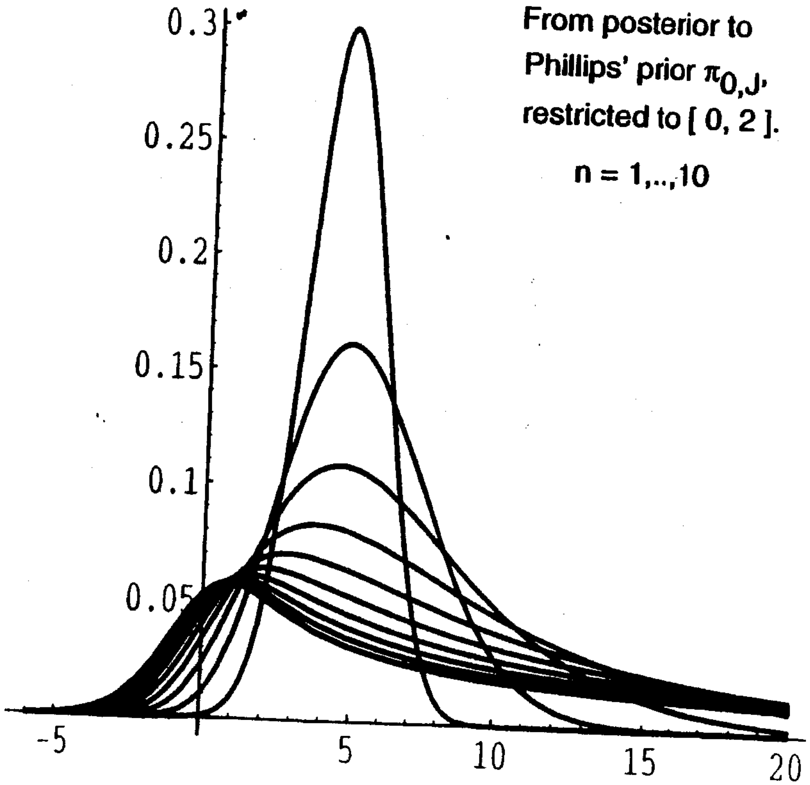


Figure 9

(Predictive densities for y_t, n steps ahead)

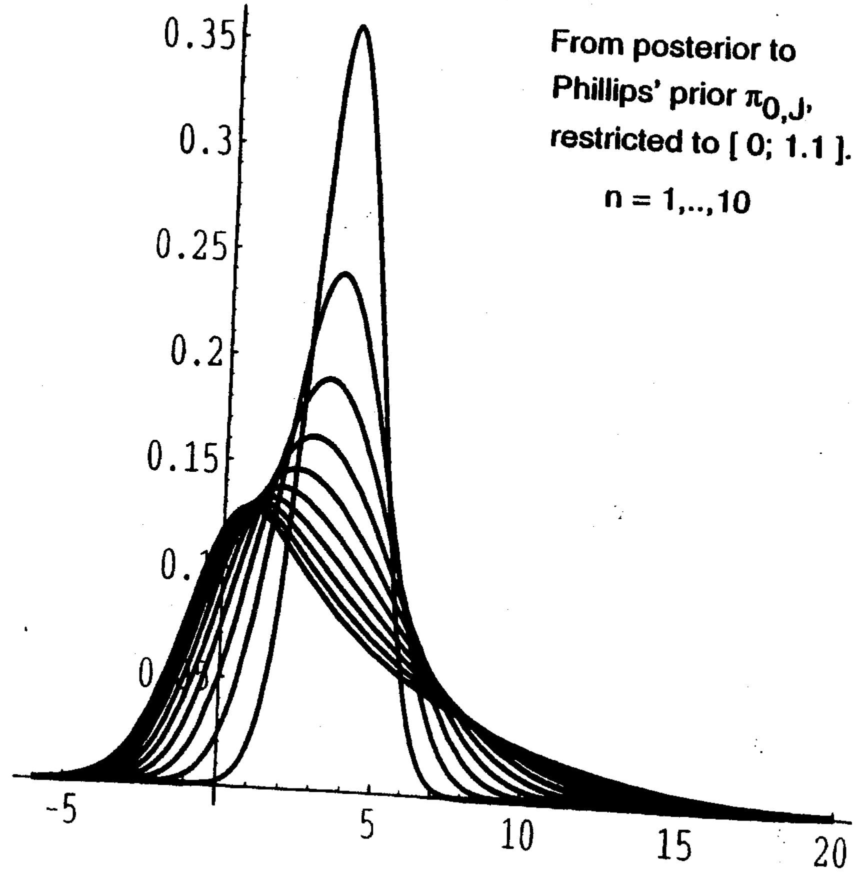
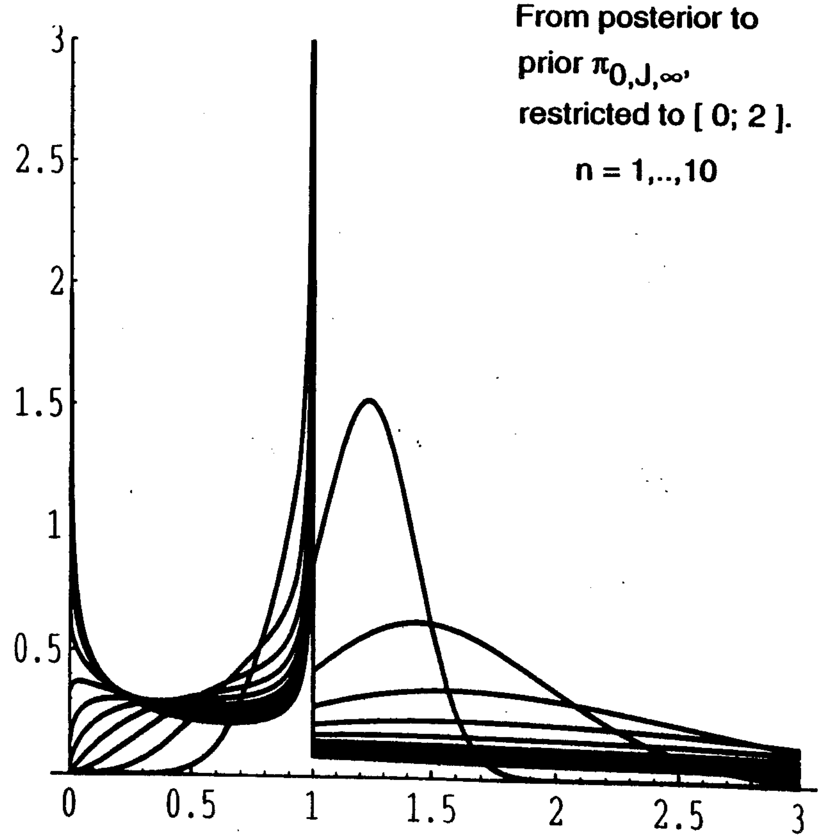


Figure 10

(Densities for the value of the impulse response function r(n), n steps ahead)



From posterior to prior $\pi_{0,J,\infty}$, restricted to [0; 2].

n = 1,..,10

Figure 11

(Predictive densities for y_t, n steps ahead)

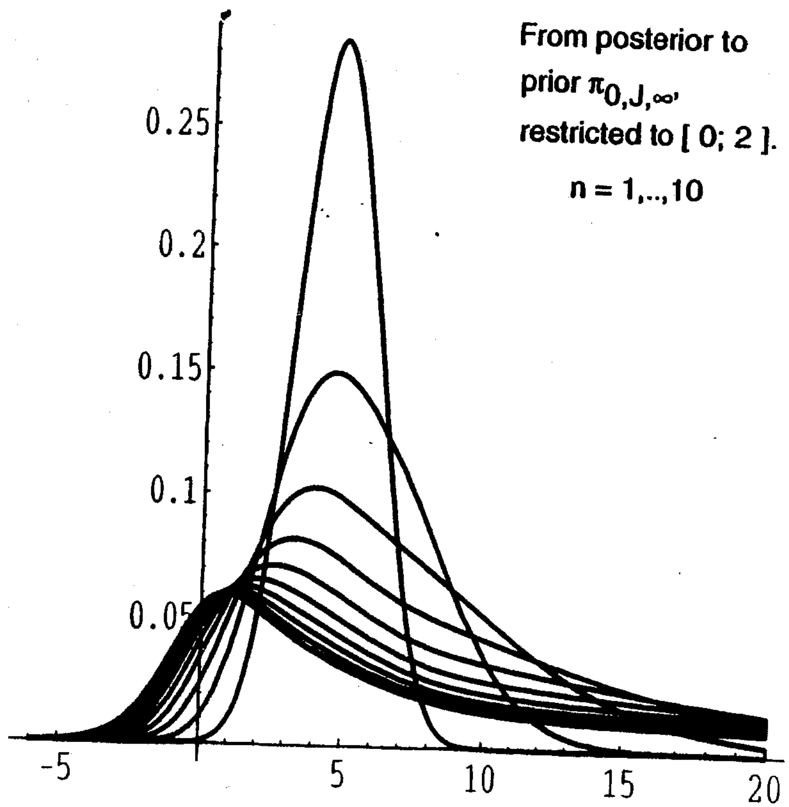


Figure 12

(Predictive densities for y_t, n steps ahead)