COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 986

BAYESIAN ROUTES AND UNIT ROOTS: DE REBUS PRIORIBUS SEMPER EST DISPUTANDUM

by

Peter C. B. Phillips

July 1991

BAYESIAN ROUTES AND UNIT ROOTS:

DE REBUS PRIORIBUS SEMPER EST DISPUTANDUM

by

P. C. B. Phillips

Cowles Foundation for Research in Economics
Yale University

novo consilio nunc opus est!

ABSTRACT

This paper provides detailed responses to the following 8 discussants of my paper "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends": Gary Koop and Mark Steel; Edward Leamer; In-Moo Kim and G. S. Maddala; Dale J. Poirier; Peter C. Schotman and Herman K. van Dijk; James H. Stock; David DeJong and Charles H. Whiteman; and Christopher Sims. This reply puts new emphasis on the call made in the earlier paper for objective Bayesian analysis in time series; it underlines the need for a new approach, especially with regard to posterior odds testing; and it draws attention to a new methodology of Bayesian analysis developed in a recent paper by Phillips-Ploberger (1991). Some new simulations that shed light on certain comments of the discussants are provided; new empirical evidence is reported with the extended Nelson-Plosser data supplied by Schotman and van Dijk; and the new Phillips-Ploberger posterior odds test is given a brief empirical illustration.

*All of the computations and graphics reported in this paper were carried out by the author in programs written in GAUSS on a ZEOS 386 20 mhz PC. My thanks go to Glena Ames for keyboarding the manuscript of this paper and to the NSF for research support under Grant No. SES 8821180.

I. INTRODUCTION

My original motivation in writing an essay "to criticize the critics" (hereafter, simply "Critics") was to confront some rather strident published criticisms of classical econometric methodology, with special reference to the problem of testing for the presence of stochastic trends. It hardly needs to be said that one does not bother to mount a critique of work that one considers worthless. Indeed, the Bayesian critiques of Sims (1988), Sims-Uhlig (1988/1991) and DeJong-Whiteman (1989a) raised issues of Bayesian doctrine and practice that were important and sorely in need of discussion. "Critics" provided a convenient vehicle for initiating that discussion. The paper also provided an opportunity to illustrate an alternative Bayesian methodology and perform a fragility analysis of the influence of different priors and lag specifications on posterior inferences with a well-used empirical data set. The discussants have now carried the debate and the empirical work further. I thank them all for their responses to "Critics," for their thoughts and reactions to my work and for their various empirical contributions. The updated Nelson-Plosser data set prepared by Schotman and Van Dijk is especially valuable in the present context and will be useful to all macroeconomists interested in historical time series.

No doubt the debate itself will help to clarify the ground that divides classical unit root econometrics and Bayesian alternatives and in so doing should help to reduce the confusion that this division seems to have spread. Of course, it is unrealistic to expect that all of the issues will be resolved. Some of them are particular to the Bayesian paradigm, like the formulation of priors, alternative routes to Bayesian inference and difficulties in constructing objective Bayesian tests of point null hypotheses such as a unit root. On such matters there is certainly plenty of room for dispute. Hence, one of the themes that I have selected for this response is carried in the title, *viz., "de rebus prioribus semper est disputandum"* (about priors there is always disputing). In short, subjective Bayesians can be expected to disagree.

The second theme of this reply puts new emphasis on the call that I made in "Critics" for objective Bayesian analysis in time series. After reading the responses to "Critics" I am now even more convinced of the need for objective methods of Bayesian time series analysis. With the sole exception of James Stock, the discussants put forward an array of competing subjective analyses that seem to please them more than the objective analysis of "Critics." Taken individually, their subjective alternatives are of some interest. Taken collectively, they demonstrate the striking fragility of subjective analysis and its consequent failure to be scientifically convincing. To me the call for objective analysis is now even more urgent. I put the matter bluntly in the header to this reply: *novo consilio nunc opus est!* (now there is a need for a new approach!).

In replying to the comments of the discussants I have the opportunity to offer some further thoughts on Bayesian inference and to draw attention to new research that I have underway with Werner Ploberger that contributes directly to the second of the themes mentioned above and responds to the charge of the header. In addition, I shall report new simulations that shed light on certain comments of the discussants and new empirical evidence with the extended Nelson-Plosser data set.

II. RESPONSES TO THE DISCUSSANTS

In order to appreciate the proper context of the discussants' remarks it will be helpful to recall the main elements of the "Critics" paper. It will be convenient to list these as numbered items for subsequent reference as follows:

II(i) A rebuttal of the Sims and Sims-Uhlig critiques that classical methods are logically unsound and Bayesian methods inherently superior;

II(ii) A critique of the use of flat priors on the coefficients as uninformative in the context of parametric time series models;

II(iii) Arguments that data conditioning principles are not innocuous in models with time series regressors;

II(iv) Construction of invariant Jeffreys priors that allow for stochastic nonstationarity and purport to represent ignorance in time series models (hereafter, "Critics"-priors);

II(v) Use of the Laplace method to reduce multidimensional posteriors, and illustrations of its use in models with time trends and multiple lags;

II(vi) Simulation exercises to evaluate the performance of flat-prior and Jeffreys-prior Bayesian analyses in models with time trends, multiple lags and various error structures when there is a stochastic trend in the data;

II(vii) A comparative Bayesian analysis of stochastic nonstationarity under flat-priors and "Critics"-priors applied to the Nelson-Plosser data set.

The discussants' comments deal almost exclusively with the issue of the priors used in "Critics" (item II(iv)), the simulation experiments (item II(vi)) and the empirical study (item II(vii)). There is little said about the other items and this is no doubt explained by a desire to join battle on the central front of the empirical methodology as it was implemented in "Critics." Nevertheless, even in that restricted context, items II(iii) and II(v) are also of great importance: II(iii) because data conditioning is a vital element in all

Bayesian methodology and, as we shall reemphasize below, has major consequences in terms of the proper interpretation of empirical results; and II(v) because it provides a convenient analytic short-cut to manageable posteriors, because it has until now not been used in Bayesian econometric exercises and because, by virtue of the huge reductions in computation time that it offers over alternative Monte Carlo integration methods, it opens the door to a proper frequentist evaluation of Bayesian methodologies.

I shall deal with the discussants' comments in the order in which they were assembled and presented to me for response.

(a) Gary Koop and Mark Steel (KS)

KS start their comments with some general issues of prior selection and conclude that the priors in "Critics" may disturb some Bayesians. This is hardly surprising. ***De rebus prioribus semper est disputandum.*** The comments of the other discussants, notably Edward Leamer and Dale Poirier, readily confirm this. KS argue that for a prior to be reasonable it must be acceptable to "a wide variety of researchers." They indicate their own subjective beliefs about what is acceptable, *viz.* all prior probability on stationary, unit root and mildly explosive models. (What about explosive models with floors and ceilings, threshold models, models with intermittent breaks and so on?) In short, they view the "Critics" priors as stacking the odds against stationary models.

Subjective reactions such as these are to be expected. One of the main classical objections to the Bayesian approach is that priors often do matter (the results of "Critics" certainly confirm this) and that if Bayesian methods are to be "scientific" (i.e. reproducible by other researchers who are working with the same model and data) then objective methods of generating priors are required. Hence, the interest in objective Bayesian analysis, priors that represent ignorance and model-based reference priors. Unfortunately, it is now apparent that, despite intense interest in this approach, no widely accepted procedure has emerged. Instead, one often encounters in practice what I characterized in "Critics," I believe quite justly, as the rather mechanical use of flat priors on the coefficients. Unfortunately in my view, it is this latter practice that has caught on in econometrics in recent years. No doubt the practice stems from convenience and from the knowledge that the procedure produces sensible results (equivalent to those of classical methods) in regression contexts with *fixed* regressors. Moreover, as it might be argued by the proponents of this practice, even in the most general context the procedure simply involves scaling the likelihood and who can say that is bad practice? After all, examination of the likelihood function is recommended practice in classical statistics and behavior of the likelihood function asymptotically determines the properties of that most revered of estimators the maximum likelihood estimator (MLE). However, to counterargue the point, classical theory recognizes that the MLE is not always a good estimator, is often badly biased in finite samples and that compensation to account for its sampling properties is always required. Bayesian analysis, because it conditions on the data, does not permit this. In effect, the likelihood function becomes a prison whose window may sometimes afford only a limited view of the parametric landscape. "Critics" showed that this is exactly what can happen in time series Bayesian analysis with flat priors, where the posterior inherits the poor sampling properties of the MLE (notably its biased location). The Jeffreys prior provides some way of compensating for this deficiency. I never claimed in "Critics" that it was perfect or even adequate. I simply argued that the mechanical use of flat priors in time series models was questionable practice, and that the Jeffreys prior often leads to a wider landscape of parametric possibilities (i.e., greater posterior uncertainty). The simulations in "Critics" document the poor sampling properties of flat prior Bayesian inferences in models with a unit root and show those based on Jeffreys priors to be superior but still biased. The upshot is that Bayesian inferences can indeed be fragile in this context, that sensitivity analysis is required and that moderation and qualification are needed in reporting empirical results obtained with these Bayesian methods. Hardly a radical conclusion and one that I am glad to see that other discussants accept!

KS are disconcerted that under the "Critics" prior the posterior is not proportional to the likelihood and that the method of generating the "Critics" prior violates the likelihood principle (LP) and makes the prior sample size (T) dependent. The "Critics" prior functionalizes the prior on the sample size and for good reason. *A priori* we know that the amount of information in an autocorrelated time series trajectory will depend on its length. It will also depend on other parameters like the sampling interval (here $h = 1$). This would naturally be accommodated in a general way if we were to embed the model in a continuous time system, which would lead to the representation $\rho = \exp(h\theta)$ where h is the sampling interval and θ is the continuous time autoregressive coefficient. Thus, the model-dependent features of the "Critics" prior can be construed as advantages.

It is easy to forget that parameters themselves are model-dependent. In a classical approach parameters are selected because they represent entities of inherent interest (like savings propensities or elasticities), because in the construction of the model it is believed that they represent quantities which are close to being fixed (i.e. inherently less variable than the data) and because they are a convenient mechanism of achieving data reduction for inferential purposes. Bayesian methods treat the parameters as random and once the likelihood is constructed sample space considerations are subsequently ignored (via the operation of the LP). Much has been written about the LP and its validity as an operating principle. Econometricians who wish to see many of the different statistical perspectives on this topic will find the collection of essays by Basu in Ghosh (1988) and the references therein a valuable source of information and debate. Unfortunately, much of the discussion on this topic does not relate to time series data. A notable exception is an important early article by Barnard, Jenkins and Winsten (1962), hereafter BJW, which is seldom referenced by econometricians. BJW argue persuasively for looking at the whole course of the likelihood function in empirical work, a message many applied econometricians will agree with. Even though BJW do not argue for a full Bayesian approach with random parameters, their suggestion of conditioning on the data and using only the likelihood function for inference meets with intense debate in the resulting discussion with contrary positions being put forward by most of the discussants, both classical and Bayesian. Many (e.g. Durbin, Whittle, Bartlett and Kendall) emphasize that sampling properties are especially important in the time series examples given by BJW and these are neglected when one's attention is restricted to the likelihood alone. Stein points out that the likelihood is not a probability density and that no Jacobian enters as a factor when coordinates in the parameter space are transformed. (One of the objects of the Jeffreys prior, of course, is to introduce the Jacobian and thereby ensure that posterior probability statements are preserved under a change of coordinates.) Even a cursory reading of this literature will persuade many that the LP is fraught with complications. Thus, if the "Critics" prior is found by KS to violate the LP than I can only say that I do not find this disturbing. In time series models, especially, I believe it is important in inference to compensate for the fact that the likelihood tells only part of the story. In a very recently completed paper, Phillips and Ploberger (1991) make a strong case for a different conceptual framework in Bayesian inference from time series that explicitly recognizes this deficiency and shows how to compensate for it in inference. I shall have occasion to refer to the results of that paper later in this reply.

KS state that the "Critics" prior impacts on the posterior "even asymptotically." Only brief orders of magnitude arguments are sketched in their comments, and their reasoning is undeveloped and unclear. I believe that they are quite wrong on this point. Let us look at the problem more fully. Take the case of the AR(1) studied in Section 3.2 of "Critics." The marginal posterior for ρ is given by equation (11) of "Critics" and, as pointed out in remark (3) following that equation, the asymptotic behavior of the posterior when $\rho_0 = 1$ depends on that of

$$I_{\rho\rho}^{1/2}\left[1 + T(\rho - \hat{\rho})^2\int_0^1 W^2\right]^{-T/2}. \tag{1}$$

(Here and elsewhere in this reply I use the same notation as that of "Critics.") Note that $\hat{\rho} = 1 + O_p(T^{-1})$ and (1) is of $O_p(T)$ for ρ in a neighborhood of $\hat{\rho}$ of $O(T^{-1})$ and of smaller order for ρ outside such a neighborhood. Thus, behavior of the posterior asymptotically is determined within such a neighborhood. Let $\rho = \hat{\rho} + h/T$, say. Then, noting that $I_{\rho\rho}^{1/2} \sim T/2^{1/2}$ and standardizing so that (1) has a proper limit as $T \to \infty$, we find the following limit behavior for the "Critics" posterior:

$$\exp\left\{-\frac{1}{2}h^2\int_0^1 W^2\right\}. \tag{2}$$

A similar argument yields identical limiting behavior for the posterior from a flat prior. Moreover, upon appropriate scaling the likelihood function also has a limit that is proportional to (2), as shown in a much more general setting in Phillips and Ploberger (1991, Theorem 5.3). Thus, the "Critics" posterior, the flat-prior posterior and the likelihood all have identical limiting behavior, contrary to the assertions of KS.

Many of the remaining criticisms in KS rely on their assertion that the "Critics" prior is not dominated by the likelihood asymptotically. Since the assertion is false I will give no further attention to those aspects of their comment.

KS make some brief final remarks about the empirical results in "Critics." They interpret those results differently from the interpretations given in the paper. They find the results to be "robust to prior selection" and the use of the "Critics" prior to lead only to "a slight increase in the probability of the explosive model holding." I can only conclude that either their criteria for evaluation are very different from mine or that they have read this section of "Critics" very quickly. I draw attention to the fact that for five series (industrial production, consumer prices, velocity, bond yields, and stock prices) the differences in the posterior probabilities of $\rho \geq 1$ are substantial in both absolute and relative terms (for four of these series the "Critics" posterior gives a value over 100% greater than that of the "flat" posterior). I see no way in which such results can be regarded as robust. Even a cursory look at the figures shows evidence of fragility in these posteriors. Would KS have to see 300%, 500% or even greater differences before they began to suspect sensitivities to the prior? Clearly such matters are subjective without the use of some agreed loss function. For myself, I think the evidence of fragility is strong for some of the series (like stock prices, velocity and industrial production). For others (like real gnp and real gnp per capita) robustness is certainly evident. Both sets of outcomes are interesting and deserve recognition. "Critics" tried to put this perspective forward in a moderate way. KS are unconvinced. Obviously, more dramatic results are needed to shift some subjective Bayesians off their own posteriors.

(b) Edward Leamer (EL)

EL's warm welcome to the small but growing circle of Bayesian econometrics enthusiasts is appreciated. In econometrics, like other realms of academe, it is important to resist the phenomenon of the "invisible college." The invisible college grows up to serve a school of like-minded researchers but ends up protecting that school in a dangerously defensive way from the very outside influences that can nurture it in new directions and thereby help it to prosper. While "Critics" was not expressly written to challenge the invisible college of flat-prior Bayesians, it was certainty written to offer new perspectives, alternative research directions and to point out the most egregious shortcomings of that school of research.

Treating "Critics" as an outside influence precisely in this intended fashion, EL has responded in a most positive way. He takes on board what he deems useful, *viz.* (i) the demonstration that priors do matter in time series models, and (ii) the evidence that uniform priors are not always a satisfactory fall-back prior. He discards what he dislikes, deems unreasonable or views as irrelevant, *viz.* (i) concern over the property of invariance, (ii) the dependence of the "Critics" prior on the sample size and (iii) issues of objectivity in Bayesian inference. He then outlines his own principles in determining a prior and suggests the major steps that are needed to characterize the way in which inferences ultimately depend on prior information.

EL's framework is useful and, if one accepts his subjective orientation, I see no reason for disagreement. (Note that his comment on sample size dependence is dealt with in II(a) & (h) and was, of course, already recognized in "Critics.") That being the case, let us see how results change under Leamer-type priors. To do the analysis I shall use the following family of modified ignorance priors (called ε-priors) based on those developed in ongoing work by Eric Zivot and myself (1991):

(3)
$$\pi_\varepsilon(\rho) = \sigma^{-1}\alpha_0(\rho)^{1/2}\exp\left\{-\rho^{2c(\varepsilon)}\right\}$$

where $\alpha_0(\rho)$ is the same function of ρ as that in "Critics" (see equation (14)) and

(4)
$$c(\varepsilon) = -1/4\varepsilon + (1/4\varepsilon)(1 + 4\varepsilon T)^{1/2}$$

is a function selected to ensure that (3) attains its modal value around $\rho = 1+\varepsilon$ after which the prior density falls away rapidly. The parameter ε may be set to correspond to the investigator's prior beliefs.

I shall employ two such priors. The first I shall call Leamer (i) and it is based on (3) above with $\varepsilon = 0.001$, i.e.

$$\text{Leamer (i)} - \text{prior} = \pi_{0.001}(\rho).$$

This prior captures Leamer's suggestion of a suitable prior for the unemployment rate series, *viz.* a prior that rises smoothly until around $\rho = 1$ and then falls away smoothly forever. It has two deficiencies in relation to Leamer's description: it does not fall discontinuously at $\rho = 1$ (before smoothly declining thereafter); and it does not start at zero at $\rho = 0$. Neither difference seems very important and they could if necessary be captured with some further modification to $\pi_{0.001}(\rho)$.

Leamer's description of his prior for the series real stock prices (= stock prices/CPI) is more difficult to capture and is not well represented by (3) as it stands. Leamer requires the prior to fall steeply on either side of $\rho = 1.03$ but "steeper to the left than to the right." To capture this shape I suggest the use of a reverse ε-prior that produces the mirror image of (3). This is achieved by defining a new function as follows:

$$\pi_\varepsilon^*(\rho) = \begin{cases} \pi_\varepsilon(2(1+\varepsilon) - \rho), & \rho \le 2(1+\varepsilon) \\ \pi_\varepsilon(2(1+\varepsilon)), & \rho > 2(1+\varepsilon). \end{cases}$$

Setting $\varepsilon = 0.03$ we have my formulation of Leamer's prior for the real stock price series, i.e.

$$\text{Leamer (ii) - prior} = \pi_{0.03}^*(\rho).$$

This prior has $\pi_{0.03}^*(\rho) = \pi_{0.03}(\rho)$ at $\rho = 1.03$ while $\pi_{0.03}^*(1.03 + x) = \pi_{0.03}(1.03 - x)$, thereby giving $\pi_{0.03}^*(\cdot)$ the mirror image form of $\pi_{0.03}(\cdot)$ around $\rho = 1.03$. In this way $\pi_{0.03}^*(\cdot)$ falls away steeper on the left of its peak than it does on the right.

Figures 0(1) and 0(ii) here

The Leamer (i) and Leamer (ii) priors are graphed against the "Critics"-prior in Figure 0(i) in levels-density form and in Figure 0(ii) in log-density form. As is apparent from these graphs, the Leamer (i) prior is very similar to the "Critics"-prior until around $\rho = 0.95$ after which its rate of increase falls and then after $\rho = 1$ the prior declines rapidly. The Leamer (ii) prior is very different in shape from both the Leamer (i)- and "Critics"-priors. It falls away very rapidly on the left from its peak around $\rho = 1$ but much less rapidly on the right, thereby capturing the right skewness of the prior that is implied in Leamer's description. We could move the peak of Leamer (ii) further to the right by changing ε in the $\pi_\varepsilon(2(1+\varepsilon) - \rho)$ representation. But this hardly seems necessary as $\pi_{0.03}^*(\rho)$ has all the essential elements in Leamer's description.

TABLE 1

Posterior Probabilities of Stochastic Nonstationarity

Model = AR(1) + trend

	$P_J(\rho\geq 1)$	$P_F(\rho\geq 1)$	$P_L(\rho\geq 1)$	$P_J(\rho\geq 0.975)$	$P_F(\rho\geq 0.975)$	$P_L(\rho\geq 0.975)$
Unemployment	0.126	0.001	0.001	0.129	0.002	0.002
Real Stock Prices	0.288	0.024	0.148	0.346	0.065	0.453

Using the AR(1) model suggested by Leamer for this analysis but allowing also for a deterministic trend, we have computed and graphed the marginal posteriors of ρ under the critics-, F- and Leamer-priors. The results for the unemployment rate series are shown in Figure 1 and those for real stock prices in Figure 2. Posterior probabilities for the sets $\{\rho \geq 1\}$ and $\{\rho \geq 0.975\}$ are given in Table 1 for the three different priors.

Figures 1 and 2 here

For the unemployment rate series the F-posterior and L-posterior are very close in shape. Neither posterior attaches any weight to the nonstationary region. Thus the Leamer (i)-prior and flat-prior give nearly identical results. There is no fragility here. Only the J-posterior attaches any weight to nonstationarity and this takes the form of a minor mode around $\rho = 1.3$. As discussed in "Critics," such posteriors lead to disjoint Bayes confidence sets and are often simply indicative of a greater uncertainty about ρ than is immediately apparent in the MLE $\hat{\rho}$ and the F-posterior which is centered on $\hat{\rho}$.

For the real stock price series the three posteriors are very different. The L-posterior is shifted significantly to the right of the F-posterior and is centered on a value of ρ close to but less than $\rho = 1$. The L-posterior gives an appreciable weight to the nonstationary and near nonstationary sets, in contrast to the F-posterior. The J-posterior is again bimodal but Bayes confidence sets would not be disjoint. From the last row of Table 1 it is clear that these three posteriors attach very different weights to the region $\{\rho \geq 1\}$. The largest is $P_J(\rho \geq 1) = 0.288$, which is nearly twice that of $P_L(\rho \geq 1) = 0.148$, while $P_F(\rho \geq 1) = 0.024$ is negligible. Here is a case where the prior density does matter and inferences may indeed be fragile in consequence.

These results corroborate the "Critics" finding that prior information matters in assessing the evidence in support of stochastic trends. Leamer's "sacrament of sensitivity analysis" deserves more converts, especially among those who already kneel at a Bayesian altar.

(c) In-Moo Kim and G. S. Maddala (KM)

KM focus on the "Critics" prior and argue that it gives too much weight to values of the autoregressive parameter ρ higher than 1 and thereby distorts sample evidence. They extend the Monte Carlo analysis of "Critics" to other values of ρ and find that for stationary models with $\rho = 0.95, 0.50$ the upward bias in the posterior is greater in magnitude for the "Critics" prior than the corresponding downward bias is for a flat prior. Since writing "Critics" I have performed some comparable simulations and found related results for the AR(1) model (but see below for some important qualifications). Note also that in Section 3.4 of "Critics" I reported some simulations for a model with a unit root and moving average errors and found a similar bias towards the unit root model when the true model had a large negative serial correlation coefficient giving a structure that was close to stationary.

Subject to the important qualifications below, the point is a good one and I am glad that KM have emphasized it. However, their simulations are distorted for reasons I shall explain and their results do not tell the whole story. I shall therefore report some further results that are of interest in assessing the importance of their findings. Let me start with some preliminary remarks before moving on to discuss the new simulations.

1. The "Critics" prior was constructed explicitly to allow for nonstationary ρ and to assist in a Bayesian analysis of evidence in support of nonstationarity. That framework may be expected to be less relevant for models with $\rho \sim 0.50$ than it is for models with $\rho \sim 1.0$. Moreover, most macroeconomic time series in levels or log levels have first order serial correlations well in excess of 0.50, so that KM's results are probably less relevant empirically as well.

2. As $\rho \to 0$ the model steadily loses its time series structure and the arguments given in "Critics" against the use of flat priors lose their force. Also the bias of the MLE $\hat{\rho}$ (and, hence, that of the F-posterior which is centered on $\hat{\rho}$) reduces in magnitude with ρ. So we must expect better performance from the F-posterior as $\rho \to 0$.

3. Regarding KM's simulation design and specific results my comments are as follows:

(i) KM's Tables 1-3 report posterior means under both the flat-prior and the "Critics"-prior. As pointed out in "Critics" in Remark 1 following equation (11), the J-posterior has no finite integer moments. KM's tables, therefore, report calculations of quantities that do not exist! Finite integrals like those given in KM can always be found by truncation of the limits. The problem is, of course, that the results are quite arbitrary and depend entirely on the chosen limits. Their posterior mean calculations and comparisons should therefore be ignored, together with the attendant discussion in their paper. Unfortunately, much of the discussion in KM does focus on the mean and mean/mode comparisons between the F-posterior and "Critics"-posterior. In consequence, the thrust of their critique concerning the bias of the "Critics"-posterior can be dismissed as groundless.

(ii) The posterior probability calculations in KM's tables do show evidence of an upward bias in the "Critics" posterior. But this evidence is much less dramatic than their (invalid) posterior mean calculations. As we shall see below, there is good reason to question the relevance of these posterior probability calculations also.

(iii) KM compute the sampling characteristics of the statistics |mean-mode| for the F-posterior and find sample mean values that are substantially different from zero. Yet the F-posterior is symmetric and has finite mean so that we should have mean = mode for every replication! Thus, no computation is required. Clearly, the nonzero values given by KM must result from their unnecessary computations and therefore give some idea of the computational or approximation errors involved in the KM simulations.

(iv) In their conclusion KM tells us that the "Critics" paper

(KM_1) *"shows that [under an ignorance prior] most US economic time series have a unit root under the Bayesian analysis as well."*

This conclusion is a plain misreading of "Critics." Even for the AR(1) + trend model, which is more generous to stochastic nonstationarity than the AR(3) + trend model, "Critics" found $P_J(\rho \geq 1) \geq 0.30$ for only seven series. Thus, even under the most liberal reading of the "Critics" results we would find evidence of stochastic nonstationarity for only one half of the time series. This does not come close to supporting the statement KM_1.

(v) Also in their conclusion, KM state that

(KM_2) *"...recent work, even along classical lines has shown that the evidence for unit roots is weak,"*

and cite work of Kwiatkowski, Phillips and Schmidt (1990) (hereafter, KPS) in support of this statement. KPS show how to test a null of trend stationarity in place of the more usual null of a unit root. The empirical work of KPS which is undertaken with the same data set as that used in "Critics" indicates that the null hypothesis of trend stationarity can be rejected at the 5% level for five series (industrial production, consumer prices, real wages, velocity and stock prices). These empirical results are, as KPS state, in accord with those of "Critics" (with the exception of the outcome for real wages) and do not, in my view at least, support the statement KM_2.

(vi) "Critics" argued that a good prior should reflect the prior knowledge we have of the AR(1) model that when the true value of ρ is larger the data will be more informative about ρ. KM object to this argument and state in their Introduction and Conclusion that

(KM_3) *"...this is a property of the likelihood function and should have nothing to do with the prior."*

Actually, it is a property of the *model* that the regressor in the AR(1) carries information about ρ and this is the source of the generic information that is employed in the "Critics" prior. Since the Bayesian paradigm relies on the LP and conditions on the lagged regressor variable the effective role of the "Critics" prior is to compensate for the delimiting effect of the LP (*viz.* the neglect of sample space considerations) by incorporating this generic information in the posterior. KM seem to have missed this point entirely.

4. The model used in KM's simulations is the AR(1). While I certainly gave attention to the AR(1) in "Critics" the main thrust of the work centered on models with trends, and models with trends and transient dynamics. As shown in "Critics," the F-posterior inherits the bias characteristics of the MLE $\hat{\rho}$. Since these are well known to be exaggerated by the presence of additional regressors like polynomial trends and further lagged regressors, we expect the bias to be worse in such cases. This is precisely what occurs and the simulation exercises in "Critics" provide ample evidence. How does this consideration affect the point about the "Critics" posterior being biased upwards in models with stationary ρ? To investigate this issue I ran further simulations, this time with the AR(1) + trend model

$$y_t = \mu + \beta t + \rho y_{t-1} + \varepsilon_t, \ \varepsilon_t \equiv \text{i.i.d. } N(0, \sigma^2) \tag{5}$$

as in equation (13) of "Critics." Setting $\beta = 0.025$, $\mu = 0$, $\sigma^2 = 1$ and sample size $T = 50$, I computed expected posterior probabilities of the same sets used in KM's Tables 1-3 for several values of ρ. The results are based on 10,000 replications and are shown in Table 2.

TABLE 2

Simulation Estimates of Expected Posterior Probabilities of ρ in AR(1) + trend (μ = 0.0, β = 0.025, σ^2 = 1, T = 50)

True value of ρ	Range	F-Posterior				"Critics"-Posterior			
		Expect.	Var.	Max.	Min.	Expect.	Var.	Max.	Min.
ρ = 0.95	$P(0.90 \le \rho \le 1.00)$	0.183	0.031	0.817	0.000	0.157	0.018	0.774	0.000
	$P(\rho \ge 0.95)$	0.101	0.018	0.986	0.000	0.350	0.059	0.999	0.000
	$P(\rho \ge 1.00)$	0.033	0.003	0.865	0.000	0.285	0.050	0.999	0.000
ρ = 0.80	$P(0.75 \le \rho \le 0.85)$	0.220	0.019	0.476	0.000	0.166	0.010	0.451	0.000
	$P(\rho \ge 0.80)$	0.231	0.049	0.996	0.000	0.428	0.065	0.999	0.000
	$P(\rho \ge 1.00)$	0.007	0.000	0.365	0.000	0.239	0.031	0.994	0.000
ρ = 0.50	$P(0.45 \le \rho \le 0.55)$	0.197	0.008	0.314	0.000	0.175	0.006	0.297	0.000
	$P(\rho \ge 0.50)$	0.340	0.070	0.998	0.000	0.404	0.082	0.999	0.000
	$P(\rho \ge 1.00)$	0.000	0.000	0.018	0.000	0.098	0.013	0.966	0.000

Table 2 tells a story that is very different from KM. Both the F-posterior and "Critics"-posterior are biased downwards but the former much more so than the latter. Thus, when the true coefficient is $\rho = 0.50$ we find:

$$E(P_F(\rho \ge 0.50)) = 0.340, E(P_J(\rho \ge 0.50)) = 0.404.$$

For $\rho = 0.80$, the corresponding figures are:

$$E(P_F(\rho \ge 0.80)) = 0.231, E(P_J(\rho \ge 0.80)) = 0.428.$$

These results show that the downward bias of the F-posterior is substantial, even for ρ as low as $\rho = 0.50$. Far from being biased upwards, the "Critics"-posterior is also biased downwards, but to a much lesser degree. These findings strongly corroborate the results of "Critics." While specific results are always model dependent, I find it encouraging that one of the major arguments in "Critics," is vindicated even in trend stationary systems, *viz.* that F-posterior inherit the poor sampling properties of the MLE, notably its bias.

(d) Dale J. Poirier (DJP)

DJP's welcome to the Bayesian journey is hearty and sincere and I thank him for it. The itinerary he describes has intellectual scenery that intrigues me and I like the scope for controversy en route. If the debate over "Critics" is a representative journey then the future beckons to the itinerant Bayesian. It is surely not fitting for peon's like myself to entertain the prospect of a subjective Bayesian passage and I am personally committed to objective modes of transport. But possibly some future Yale students will qualify for these tantalizing upgrades that DJP dangles before us.

DJP correctly interprets the central message of "Critics" and the spirit in which it was written. He recognizes the two key issues that affect Bayesian reasoning, *viz.*

(i) *res priores*; and

(ii) data conditioning principles.

Decisions that are made on these issues can and do affect inferences as we saw in "Critics" and again in the simulations and empirical exercises of II(a) and II(b) above. "Critics" gave its main attention to *res priores*. In later work with Werner Ploberger we have been studying the implications of data conditioning. Phillips and Ploberger (1991) examines the effects of data conditioning principles on inference in terms of the probability measures that are implied by the operation of Bayes theorem. In the context of time series data, the implications are of great importance. They ensure that the Bayesian analyst lives on a particular trajectory, as DJP and KS emphasize. The implied Bayes model (in Phillips-Ploberger terminology) is a time varying parameter model and the Bayes model measure is a conditional predictive measure, giving the predictive probability measure of the next period observation given the past history of the data *and* the MLE of the model's coefficients that this data implies. Phillips-Ploberger seems to be the first general treatment of this feature of Bayesian analysis, providing a new "frame of reference" for thinking about Bayes methods. The Phillips-Ploberger analysis shows the following:

(i) Bayes models and classical models differ in fundamentally important ways, including their probability measures.

(ii) Bayes model likelihood ratio tests can be constructed to evaluate the evidence in support of a Bayes model against an alternative reference measure, including alternative Bayes models.

(iii) Posterior odds analyses are also possible and enable an investigator to compare different Bayes models in the light of the observed data and against a background of prior odds.

The upshot of this analysis is that rigorous attention to the data conditioning principles of Bayesian analysis in time series settings forces one into a new frame of reference that delivers a new geometry of inference. When DJP enjoins me as follows,

(DJP_1) *"In short, I recommend Phillips worry about the data he sees along the road he travels and not worry about the data he might see along roads he does not travel,"*

he is appealing, in effect, for the use of the Phillips-Ploberger geometry. If one intends to follow the Bayesian route then I wholeheartedly agree. But in accepting that geometry one must also accept the reference measure that the new coordinates imply. As Phillips-Ploberger (1991) shows, the cost of living on a particular trajectory leads to a measure in which there is explicit compensation (i.e. weighting) for the chosen parameterization. In the construction of a posterior odds test of $\rho = 1$ in the AR(1) the compensator is the square root of the martingale conditional variance, i.e. $(\Sigma_1^T y_{t-1}^2)^{1/2}$ which is a form of conditional Jeffreys' prior suited to this time series example. This compensation has a major effect on posterior odds inferences. I shall briefly illustrate the use of the new criterion in the next subsection, II(d).

If, as DJP argues,

(DJP_2) *"...conditioning is at the heart of most differences over the appropriate way to conduct statistical inference,"*

then the Phillips-Ploberger geometry helps to reconcile those differences because it makes explicit the reference measures associated with Bayes and classical models. This geometry also bears on the "conditional frequentist" approach suggested in the Hinkley (1983) paper cited by DJP. By conditioning the information content of a trajectory of data at a fixed level the "frequentist" can avoid many of the inferential difficulties of classical procedures, ending up with nice Gaussian sampling distributions, as pointed out in "Critics." But there is a cost that is induced by the latent sampling variability of the compensator and this too needs to be factored into the accounting. More work on this issue with Werner Ploberger is presently in the pipeline.

DJP draws attention to concerns in the literature about the use of *J*-priors and reference priors in general. Regarding the "Critics" prior he points to its dependence on the likelihood, the sample size and the sample space mathematical expectation. Unlike other discussants (especially KM), DJP discusses the issue of dependence on the form of the likelihood, noting that other priors in regular use, like conjugate priors, have similar dependencies. But he does object to sample size dependence. I have already partially responded to this point in II(a), but more needs to be said here. DJP gives the following illustration of incompatibility that he argues the "Critics"-prior induces:

(DJP$_3$) *"Phillips' prior (9) depends on (b) [= the sample size]. Hence, in the face of T observations, using prior (9) based only on the first m (0 < m < T) observations and then using the resulting posterior as the prior for the next T−m observations gives a different posterior than arises from using prior (9) based on all T observations. Why should the way the data are processed affect the answer?"*

In fact, there is no such incompatibility. To explain why, take the case where $\sigma^2 = 1$ is known and let

$$\pi^T(\rho) = \alpha_0(\rho)^{1/2} = \text{"Critics"-prior}$$

based on a sample of T observations. The corresponding posterior is then proportional to

$$\Pi^T(\rho) = \pi^T(\rho)\mathrm{pdf}(y_1^T|\rho, y_0),$$

where $y_1^T = (y_1, y_2, ..., y_T)$ and $\mathrm{pdf}(\cdot|\rho, y_0)$ is the conditional density given ρ and y_0 (i.e. the likelihood). We now decompose $\Pi^T(\rho)$ as follows:

$$\begin{aligned}\Pi^T(\rho) &= \left[\pi^m(\rho)\mathrm{pdf}(y_1^m|\rho, y_0)\right]\left[\frac{\pi^T(\rho)}{\pi^m(\rho)}\mathrm{pdf}(y_{m+1}^T|\rho, y_0^m)\right] \\ &= \left[\Pi^m(\rho)\pi^T(\rho)/\pi^m(\rho)\right]\mathrm{pdf}(y_{m+1}^T|\rho, y_0^m) \\ &= \tilde{\Pi}^m(\rho)\mathrm{pdf}(y_{m+1}^T|\rho, y_0^m).\end{aligned}$$

Here $\Pi^m(\rho) = \pi^m(\rho)\mathrm{pdf}(y_1^m|\rho, y_0)$ is the posterior based on the first block of data y_1^m and the "Critics"-prior $\pi^m(\rho)$ for the sample of m observations. We now use this posterior as the prior for the next Bayesian analysis with the second block of data y_{m+1}^T. However, recognizing that the second block has $T-m$ observations (giving a *total* sample size of T consecutive observations) we scale the new prior $\Pi^m(\rho)$ by the factor $\pi^T(\rho)/\pi^m(\rho)$ to adjust for the expected information content of T observations relative to that of the m observations upon which $\Pi^m(\rho)$ is based. Note that under a flat-prior the ratio $\pi^T(\rho)/\pi^m(\rho)$ would be a constant and $\Pi^m(\rho)$ would equal $\tilde{\Pi}^m(\rho)$. Under the "Critics"-prior it is recognized that the information content of data is different between a consecutive sequence of T observations and two independent sequences that sum to T. Again, the "Critics"-prior takes the model and time series nature of the data into account whereas the flat-prior does not.

In cases where the number of observations T is random and ancillary, like DJP's coin toss that determines whether there are $T = 100$ or $T = 50$ observations, we could indeed condition on T as its distribution contains no information about ρ. The conditional expected amount of information in the data about ρ would then depend on the realized value of T. Such an approach would be in keeping with more traditional Bayesian thinking (c.f. also Barnard Jenkins and Winsten's (1962) treatment of the LP). The likelihood would not depend on the sample stopping rule for T, just its actual realization. Likewise the prior would be conditional and represent the conditional anticipated amount of information in the data. We could also, of course, build the sampling principle for T into the prior and it would then reflect the average anticipated amount of information in the data. Then for trajectories with $T = 50$, the prior would overcompensate in the explosive range and correspondingly undercompensate when $T = 100$. Note that the Phillips-Ploberger (1991) posterior odds criterion would not suffer this ambiguity because it is trajectory based and would therefore be conditional on the outcome for T.

DJP gives an interesting case study of conventional posterior odds analysis applied to case (a) of Table 1 of "Critics." He attaches a point prior mass of 1/2 at $\rho = 1$ as hypothesis H_1 and a $N(1, \tau^2)$ prior density as H_2 with $P(H_2) = 1/2$. Thus, prior odds are equal and DJP's Table 1* reports the posterior probability of H_1 for various values of the prior standard deviation τ. For large values of τ (corresponding to a "noninformative" alternative H_2) the data clearly support H_1 and DJP rightly concludes that

(DJP_4) *"Assigning a point mass to $\rho = 1$ has a big impact."*

Let us now consider the use of the Phillips-Ploberger (1991) posterior odds test of H_1: $\rho = 1$ against $H_2 : \rho \neq 1$. Unlike DJP we do not need to set up an arbitrary prior for H_2. Indeed, H_2 in the Phillips-Ploberger geometry is the Bayes model:

$$H_\rho : y_{n+1} = \hat{\rho}_n y_n + u_{n+1}$$

where $\hat{\rho}_n = \Sigma_1^n y_t y_{t-1} / \Sigma_1^n y_{t-1}^2$ is the classical MLE of ρ. The Bayes model measure of B is denoted Q_n^B in Phillips-Ploberger notation and the Bayes model posterior odds criterion is based on the likelihood ratio dQ_n^B/dP_n which is the Radon-Nikodym derivative of Q_n^B with respect to the reference measure P_n for the random walk that applies under H_1. The decision rule is then

$$\text{if } \frac{dQ_n^B}{dP_n} > \frac{\pi_1}{\pi_\rho} \text{ decide in favor of } H_\rho$$

$$\text{if } \frac{dQ_n^B}{dP_n} < \frac{\pi_1}{\pi_\rho} \text{ decide in favor of } H_1$$

where π_1/π_ρ is the prior odds ratio (see equation (71) of Phillips-Ploberger). As in DJP, set $\pi_1/\pi_\rho = 1$. Using equation (70) of Phillips-Ploberger and the data of Table 1 of "Critics" we obtain by an easy calculation

$$\frac{dQ_n^B}{dP_n} = \exp\left\{\frac{1}{2}(1 - 0.804)^2)78\right\}/78^{1/2} = 0.506 < 1$$

thereby conclusively deciding in favor of H_1, or $\rho = 1$. Recall from Table 1 of "Critics" that in this case the flat prior gave a posterior probability of $P_F(\rho \geq 1.0) = 0.02$. Clearly DJP, Phillips-Ploberger and "Critics" all dominate the flat prior approach in this case!

(e) Peter C. Schotman and Herman K. van Dijk (SVD)

SVD present a thoughtful and productive commentary on "Critics." I thank them for these comments and I am sure the profession will join me in thanking them for the extended Nelson-Plosser data set.

Like the earlier discussants their comments focus on *res priores*. But they offer us new thoughts on the effects that different model specifications and parameterizations have on invariant priors, they argue a case for the use of a posterior odds test of the sharp null hypothesis $\rho = 1$ and they present numerical results with these methods for the original and extended Nelson-Plosser data sets.

As pointed out in "Critics," the J-posterior is biased downwards in models with a fitted intercept and trend in the sense that $E\{P_J(\rho \geq 1)\}$ is substantially lower for that model than it is for the simple AR(1). SVD notice this bias and start their comments with an investigation of the interaction between intercept, trend and autoregressive coefficients in F- and J-priors and their associated posteriors. They choose to work with the components model

(6) $$y_t = \gamma + \delta t + u_t,\ u_t = \rho u_{t-1} + \varepsilon_t,\ \varepsilon_t \equiv \text{iid } N(0, \sigma^2)$$

for this analysis. As discussed recently in Schmidt and Phillips (1989) this model formulation has the advantage that it has no surplus parameters (or variables) under the null hypothesis $\rho = 1$ and thereby is a convenient vehicle for the construction of an LM test of this hypothesis.

Note that (6) can, of course, be written in the earlier form (5) i.e.

(5) $$y_t = \mu + \beta t + \rho y_{t-1} + \varepsilon_t,\ \varepsilon_t \equiv \text{iid } N(0, \sigma^2)$$

with the explicit parameterization

(7) $$\mu = \gamma(1-\rho) + \delta\rho,\ \beta = \delta(1-\rho).$$

In "Critics" I used the following prior for the parameters of (5):

$$\pi(\rho, \sigma, \mu, \beta) \propto \sigma^{-3}\{\alpha_0(\rho) + \alpha_1(\rho, \mu, \beta)/\sigma^2\}^{1/2}$$

and showed that this prior leads to results that are generally well approximated by using the simpler expression

(8) $$\pi(\rho,\ \sigma,\ \mu,\ \beta) \propto \sigma^{-3}\alpha_0(\rho)^{1/2},$$

which implies a flat prior for μ and β. Next note that for the parameterization given in (7) and a flat prior on (μ,β) we have the following implied prior for (γ,δ):

(9) $$\pi(\gamma,\delta) = \pi(\mu,\beta)|\partial(\mu,\beta)/\partial(\gamma,\delta)| = (1-\rho)^2.$$

Combining (8) and (9) we deduce the implied prior for the parameters of (6), *viz.*

(10) $$\pi(\rho,\sigma,\gamma,\delta) \propto \sigma^{-3}(1-\rho)^2\alpha_0(\rho)^{1/2}.$$

This is a simpler and more revealing derivation of SVD's equation (4). As is apparent from (10), the prior density is degenerate at $\rho = 1$. The above derivation shows that this is a consequence of the degeneracy in the prior (9) for (γ,δ) that is implied by the parameterization (7). The degeneracy in $\pi(\gamma,\delta)$ at $\rho = 1$ is explained by the fact that when $\rho = 1$ the model is simply

(11) $$\Delta y_t = \delta + \varepsilon_t$$

i.e. γ does not occur at all and there is no surplus parameter under the null. Thus, for the model (6) we have only three parameters ($\rho = 1$, δ, σ^2) under the null $\rho = 1$ but four parameters (ρ, γ, δ, σ^2) under the alternative $\rho \neq 1$. It follows that the degeneracy in the Jeffreys prior (10) at $\rho = 1$ that is emphasized by SVD is a simple consequence of the attempt to force a four parameter model, i.e. (5), to accept a parameterization that is degenerate at $\rho = 1$.

By contrast, suppose we insist on dealing with the components model (6) and also insist on a flat prior for $(\gamma,\ \delta,\ \rho,\ \ln(\sigma))$. Since the data generating mechanism (i.e. the reduced form) is (5), with the explicit form (6) requiring the parameterization (7), we deduce that a flat prior for (γ,δ) necessitates the prior for (μ,β) of the form

$$\pi(\mu,\beta) \propto (1-\rho)^{-2}$$

which involves a singularity at $\rho = 1$.

None of these apparent "pathologies" is anything other than a consequence of forcing a degenerate parameterization, i.e. (7), on the data generating mechanism (5). The reason this parameterization (i.e. the components representation (6)) works so well in the classical analysis of Schmidt-Phillips (1989) is that the model is estimated under the null and only the efficient score is used in constructing the LM test. A

Bayesian analysis requires that we define a prior distribution for the full parameter space. If we insist, as SVD do, on working with a structural model like the components representation (6) then there will always be a degeneracy of some form due to the mapping to the reduced form (5). The pathologies they discuss are the consequence of this mapping and are not inherent weaknesses in any of the prior densities. Similar problems can be expected in any Bayesian analysis of structural models.

The second concern of SVD is to mount explicit Bayesian tests of the sharp null hypothesis $H_0 : \rho = 1$. They rightly argue that this null is of sufficient importance to warrant an explicit Bayesian test and correctly point out the limitations of the analysis in "Critics," which concerned itself with stochastic nonstationary sets (*viz.* $\{\rho \geq 1\}$) against stationary sets (*viz.* $\{|\rho| < 1\}$). SVD encounter two problems in setting up a Bayesian test of H_0. First, as DJP remarked in his commentary, there is no general objective procedure for mounting posterior odds tests of sharp null hypotheses presently available in the literature. SVD also tell us this in their conclusion, saying that

(SVD_1) *"Objective Bayesian methods for testing a sharp null hypothesis do not exist"*

and they refer us to Berger and Delampady (1987). This is therefore a good point at which to draw the reader's attention to new methodology in the recent paper by Phillips and Ploberger (1991). The Phillips-Ploberger (1991) Bayes model posterior odds test is objective, allows one to test a sharp null hypothesis such as H_0, is available for use in the present context and, indeed will be briefly illustrated below. Phillips-Ploberger is a direct response to the call *"novo consilio nunc opus est"* in the header to this paper.

Second, SVD encounter difficulty in the computation of the posterior probability of H_0 because the surplus parameter (γ) under the null $\rho = 1$ in the components model leads to a divergent integral. SVD overcome this difficulty by using a normal prior on γ of the form $N(y_0, \sigma^2/(1-\rho^2))$. With this prior on γ the posterior probability of $\rho = 1$ *is* finite and can be computed. The posterior odds calculation then proceeds as usual once a prior for ρ on the stationary set is specified.

In dealing with these difficulties SVD follow the treatment (of the same problem) given in their earlier work (1991) on real exchange rates. While applauding their efforts to resolve the difficulties of the usual approach in this case, I am troubled by the rather arbitrary choices that are involved. First the prior on γ is one of convenience. It has some relevance for stationary models, but is unjustified under the null. Since the null is where it is critically needed, this seems most unfortunate. Second, the prior on the alternative stationary set $[\alpha,1)$ is arbitrary. Any choice is possible. SVD seem to follow their earlier work (1991) and use a uniform prior on $[\alpha,1)$ but I could not find this clearly stated in their paper. Also arbitrary is the choice of α, which will certainly affect the posterior probability of stationarity and the odds ratio. These arbitrary elements go beyond the setting of prior odds and, at least to me, are disturbing. No doubt, many subjective Bayesians would be quite content to proceed with intelligent and reasonable choices of α and the priors on γ and ρ. But how do we assess what is reasonable? To give but one example SVD's choice of α is $\alpha = 0.8$. No doubt this depends on the time unit or sampling interval, which is one year for the Nelson-Plosser data. But should the choice be the same for stock prices, money and gnp? Is a sensitivity analysis warranted? Should we consider alternative priors? And so on. The posterior probability results in SVD's Table 1 also illustrate the problem. Against what values do we calibrate the posterior probabilities of $\rho = 1$ given in the final column. Some of the results like those for unemployment ($P(\rho = 1) = 0.11$) and stock prices ($P(\rho = 1) = 0.968$) lead to clear decisions. Others, like employment where $P(\rho = 1) = 0.614$ are less conclusive. Some objective correlative would be useful here but is not provided. While the posterior probability calculations in "Critics" suffer to some extent from a similar weakness, the situation is more critical here because an odds ratio test is being employed. The point of a posterior odds test is to perform a test and for this we need a clear criterion. When there are too many arbitrary choices in the construction of the posterior odds and no clear decision criterion then it strains credibility to say there is a test. Herein lies the rub and the reason why there is a need for an objective basis for Bayesian tests of a sharp null hypothesis.

Now let us consider the Phillips-Ploberger approach. All of the basic ideas and methods are laid out in the recent Phillips and Ploberger (1991) paper. A full development and application of the methods is presently in progress and will appear elsewhere. I shall give a brief illustration to whet the reader's appetite here. Suppose we allow for a general time series model with trends and multiple lags, i.e.

$$\Delta y_t = h y_{t-1} + \Sigma_1^k \varphi_i \Delta y_{t-i} + \mu + \beta t + \varepsilon_t, \tag{12}$$

with $\varepsilon_t \equiv$ iid $N(0, \sigma^2)$. We wish to employ a posterior odds test of $h = 0$, i.e. a unit root in (12). In Phillips-Ploberger this is achieved by using the Radon-Nikodym derivative of the Bayes model measure Q_n^B (*i.e.* h unrestricted) with respect to $Q_n^{B_0}$ (i.e. h restricted to $h = 0$). Calculations in Phillips and Ploberger (1991b) show that

$$\frac{dQ_n^{B_h}}{dQ_n^{B_0}} = \frac{\exp\left\{(1/2)\hat{\sigma}_n^{-2}\hat{h}_n^2 y_{-1}'Q_x y_{-1}\right\}}{\left\{y_{-1}'Q_x y_{-1}/\hat{\sigma}_n^2\right\}^{1/2}} . \tag{13}$$

Our decision rule with equal prior odds is then simply

(14) $$\text{if } \frac{dQ_n^{B_h}}{dQ_n^{B_0}} < 1 \text{ accept the hypothesis } H_0 : h = 0 \text{ in (12).}$$

Table 3 shows the results of this objective posterior odds test applied to the Nelson-Plosser and extended Nelson-Plosser data sets. The tabulated results refer to model (12) with $k = 3$ lags, as in the original Nelson-Plosser study. The outcome of the Phillips-Ploberger test is quite decisive for both data sets. For the 1970-sample we accept the presence of a unit root in the following 7 series (asterisked in the table): nominal gnp, gnp deflator, CPI, nominal wages, velocity, bond yields, stock prices. For the 1988-sample we accept the presence of a unit root in 8 series (again asterisked in the table): these are the same series with a

TABLE 3

Phillips-Ploberger Bayes Model
Posterior Odds Unit Root Tests

Series	Posterior Odds = $dQ_n^{B_h}/dQ_n^{B_0}$	
	1970 Sample	1988 Sample
Real gnp	2.9846	8.1780
Nominal gnp	0.2945*	0.1375*
Real gnp per capita	4.0038	12.6481
Industrial production	14.1071	18.5256
Employment	10.2765	18.1569
Unemployment rate	64.6225	223.3859
Gnp deflator	0.8465*	0.0831*
CPI	0.1306*	0.0337*
Nominal wages	0.5970*	0.3258*
Real Wages	3.0919	0.1330*
Money stock	2.7962	1.3877
Velocity	0.1232*	0.0705*
Bond yields	0.0463*	0.2879*
Stock Price	0.5365*	0.2662*
Decision rule: reject unit root if $dQ_n^{B_h}/dQ_n^{B_0} > 1$; number of lags $k = 3$		

unit root accepted for the 1970-sample plus real wages. If we use a criterion of "accept a unit root if $P(\rho = 1) \geq 0.75$" in SVD's Table 1 then the SVD results for the 1988-sample agree precisely with those of the Phillips-Ploberger test. Table 3 also shows that there is a strong rejection of the presence of a unit root for 5 series *viz.* real gnp, real gnp per capita, industrial production, employment and the unemployment rate; and there is a marginal rejection of a unit root for the money stock series. Note also that the strongest rejection is for the unemployment rate series and that, with the exception of the money stock series, the rejections are stronger for the 1988-sample than the 1970-sample. Furthermore, with one exception (bond yields) the evidence in support of the presence of a unit root in the series is also stronger with the 1988-sample. Thus, our results corroborate SVD's finding that the extra data have a noticeable impact. Finally, we observe that the only decision rule change between the 1970-sample and 1988-sample is the decision to accept the presence of a unit root in the real wage series.

(f) James H. Stock (JHS)

It is a pleasure to see a fellow classical among the discussants and I welcome Jim Stock's comments. JHS expresses broad agreement with the central message of "Critics" and it is apparent from his comments that we see the material issues of both Bayesian and classical modes of inference from a very similar perspective. His discussion of the difficulties with the Bayesian approach is characteristically lucid and informative. His views on the likelihood principle come close to my own (ref. my earlier response to KS) and I strongly endorse his position that the classical alternative is a viable and productive one.

Since there is so much common ground between us, I can keep my reply brief. There are two points that deserve discussion.

1. The conventional view is that it is necessary to declare one's inferential philosophy at the outset. That is, either one adopts a Bayesian or a classical approach to inference. Sims (1988) and Sims-Uhlig (1988/1991) tried to use the big differences between Bayesian and classical methods in nonstationary models to force this choice into the open and suggest that the classical approach was in some way logically unsound. While certainly not accepting the later proposition, JHS has given some ground to this view and he clearly feels that the issue of choice is inevitable. Thus,

(JHS_1) *"Second, as Sims (1988) and Sims and Uhlig (1988) emphasized, because of this discrepancy between the Bayesian and classical results, researchers must take a stand on whether they are classical or Bayesian statisticians."*

I believe that it is far too pessimistic to put the choice in those stark terms. Bayesians, and now it seems econometric Bayesians, often try to force a choice in this way and then, like Sims (1988), try to show that there is something very inadequate in the classical apparatus. One of the reasons we use the term "classical methods" is that they are truly classic in the sense that they have endured while other approaches have fallen by the wayside. Note that I resist the terminology "frequentist" and other terms like "Berkeleyan" (Lindley, 1990) even more strongly.

In time series applications, the ideas that underlie classical inference seem closer to those of the likelihood principle and consequently Bayesian inference. After all, in time series we are always confronted with a given trajectory and much asymptotic theory in this case (such as ergodic theory) is concerned with how much we can hope to learn if we lived on a given trajectory indefinitely rather than had the opportunity to sample alternative histories. Of course, the critical difference comes when in classical inference we admit that alternative histories are possible, inducing the essential notion of "variability" that distinguishes sampling theory. By contrast, in the Bayesian world we live forever on the given trajectory that we condition on in setting up the likelihood. However, having said this, it is possible to work out a Bayesian inferential framework with the usual tools of probability still alive and well. The trick is to work conditional on the data of the trajectory up to the present observation. The apparatus of semimartingales enables us to do this in a rather general way. Following this general idea, Phillips and Ploberger (1991) show that it is possible to bring together in a cohesive way the classical and Bayesian approaches. Interestingly an entirely new approach to testing emerges from this cohesion, one that has the objective elements of classical theory and the "decision making" apparatus of Bayes theory. This is a topic in which Werner Ploberger and I are presently do much more research.

2. I was interested in JHS's construction of asymptotic confidence intervals using local nonstationary processes and his empirical application. I have not yet seen Stock (1990) but I can understand the general proposition that the ADF t-ratio has asymptotics that rely only on the localizing constant "c" in JHS notation. The result here should parallel that of the Z_t test given in Phillips and Perron (1988) and my earlier work in Phillips (1987b) where this originated. Now JHS makes the interesting suggestion of constructing classical confidence intervals by inverting the t-ratio interval to find the corresponding interval for "c." I have one difficulty with this idea. The localizing constant "c" cannot be consistently estimated. For instance, using the diffusion process notation of Appendix A of "Critics" we have (cf. Phillips and Perron, 1988, Theorem 3):

$$T(\hat{\rho}-1) \to c\left(\int_0^1 J_c^2\right)^{-1}\left(\int_0^1 J_c\,dW\right)$$

or

$$\hat{\tau},\, Z(t) \to c\left(\int_0^1 \underline{J}_c^2\right)^{1/2} + \left(\int_0^1 \underline{J}_c^2\right)^{-1/2}\left(\int_0^1 \underline{J}_c\, dW\right)$$

where $\underline{J}_c$ is the L_2 projection residual of J_c on a constant and linear trend. While the confidence sets for c can be constructed from the second set of asymptotics above for given $\hat{\tau}$ or $Z(t)$, the first set of asymptotics for $\hat{\rho}$ show that c is unidentified in the limit. Unless I have missed the point here, this means that confidence sets for "c" produced in the JHS manner will not contract as $T \to \infty$, i.e. they do not have good asymptotic properties. We may therefore expect the confidence sets to be rather wide for finite T. This does indeed seem to be the case in the empirical application discussed by JHS.

(g) David DeJong and Charles Whiteman (DJW)

> *"The first casualty when war comes is truth"* (Hiram Johnson: speech to the US Senate, 1917).

g.1. DJW declare war

In their earlier work (DeJong and Whiteman (1989) -- hereafter, DJW^1), DJW joined Sims (1988) in strident criticism of classical analyses of stochastic trends, advocated a Monte-Carlo based Bayesian methodology they saw as superior and implemented that methodology with strongly worded conclusions that summarily turned around the empirical results of Nelson and Plosser (1982). Their comments on "Critics," which are entitled "The case for trend stationarity is stronger than we thought," (hereafter, DJW^2), put them again on the offensive. Where earlier discussants see the perspectives of "Critics" to be useful, informative and stimulating, DJW see "Critics" as simply "an attack on their work." Where other commentators correctly interpret its central message, DJW choose to ignore that message and dismiss its results. No doubt DJW are eager to defend the ground they believe they won in DJW^1. But the combative posture and military language of DJW^2 seem to invite further engagement. DJW clearly want to wage war.

Unfortunately, the first casualty of war is truth. Not only do DJW seriously overreact to a very small part of "Critics" (only three paragraphs of Section 4 of "Critics" and one paragraph of its introduction actually deal with DJW^1!) but in doing so they produce a litany of false citation, imputation and allegation. To wit:

(i) DJW assert

(DJW_1) *"He claims to adopt an "unbiased" procedure (featuring an "ignorance prior...."*

Yet nowhere in "Critics" do I make such a claim or anything that remotely approaches it and the citation "unbiased" is demonstrably false. On the contrary, in fact, I repeatedly point out that the procedure I suggest itself suffers from *bias*. The downward bias of the "Critics"-posterior is well illustrated *and* discussed in the simulation sections of the paper. Other commentators (SVD, for example), have noted it and remarked on it. And in the conclusion of "Critics," I went to some effort to emphasize the ambiguities and limitations of the objective Bayesian analysis that the paper presented (see $PCBP_4$ quoted below). The evidence, therefore, strongly rejects the validity of (DJW_1).

(ii) DJW repeatedly claim that my description of their prior is incorrect. For instance they state:

(DJW_2) *"Phillips refers to our prior as truncated and flat. Here we argue that truncation is irrelevant and flatness untrue...."*

Yet DJW^1 does employ a flat prior *on the autoregressive coefficients*! The latter qualification in italics is important and is mentioned on every occasion that I discuss their prior. In Section 4 of "Critics" I go to considerable lengths to emphasize this fact. In spite of their repeated claims like DJW_2 about the invalidity of my description DJW never attempt to quote me directly. And for good reason. I well understand and repeatedly state in "Critics" that their implied prior on the largest root Λ is not flat when the lag length $k > 1$. The point of this evasion by DJW eludes me. If they wish to avoid the criticisms I level against the mechanical use of flat priors, they have not succeeded.

(iii) DJW twice assert, unjustly, that "Critics" ignores the cost of approximating the Jeffreys prior by the prior that is actually used in "Critics" for the model with trend and transient dynamics. I shall deal with this allegation (see DJW_8 below) at a more appropriate point in the technical discussion later on.

DJW retreat into silence on many of the key issues raised in "Critics." With reference to the central items listed earlier in this section, they have nothing to say about II(i), II(ii), II(iii) and II(vi). Most surprising is their failure to address II(iii). DJW^1 leans heavily on the likelihood principle (LP) for its methodology and, indeed, LP figures prominently in its title. Yet DJW^1 makes no mention of the large body of divided opinion of the relevance and applicability of LP, more especially in the time series setting in which they unquestioningly recommend its use. DJW^2 adopts a similar silent posture. I have said much on this topic already and given some relevant references in my earlier responses so I shall say no more of it here. Economists who read DJW^1 alone will unfortunately be exposed to only one very limited perspective on the topic.

g.2. Technical engagement

DJW join battle with "Critics" on four fronts. Here engagement is rejoined on each front in turn.

(i) *Issues of specification*

In "Critics" I illustrated the effects of model specification on inference by giving posterior distributions of ρ in the general model (12) for two different choices of k ($k = 1, 3$). As I emphasized in "Critics," the analytic methods that I employed made it easy and convenient to compute posteriors for different empirical model specifications. The choices $k = 1$ and $k = 3$ seemed to cover the leading cases of interest and helped to achieve comparability with earlier work, as indeed did the model specification (12). Other choices could easily have been made, but as we shall show below they would generally have been *less* relevant. One of the objectives of "Critics" was to encourage such sensitivity analysis by showing how easily it could be done (item II(v)). It seems that DJW do not like this message but prefer the rigidities of a sharp prior model specification. At a more general level, Bayes model selection procedures could be employed for discriminatory and inferential purposes. I even indicated in "Critics" how this might proceed. DJW choose to ignore this message also.

Both the simulations and the empirical results of "Critics" show that there is often substantial sensitivity to model specification. The simulations went further than specifications of the form (12) and considered models with moving average errors in which the sensitivities to specification are known to be strong in classical methods (cf. Schwert (1987) and Phillips-Perron (1988)). I believe such exercises are very important. DJW^1 gave no hint of possible fragility to model specification and, now that the evidence has been presented, DJW^2 prefer to ignore it.

DJW do make a strong objection to my use of $k = 1$. But here their objection is on very weak ground. As indicated above, DJW^1 performed no sensitivity analysis with respect to lag length and used $k = 3$ throughout, so that readers have no grounds for believing the results in DJW^1 are robust. On the contrary, the results with flat priors in "Critics" make the fragility to lag specification all too clear. To support their earlier use of $k = 3$, DJW^2 now estimate (12) by OLS with $k = 1$ and compute the first order serial correlation coefficient of the residuals from their regression. They tell us

(DJW_3) *With few exceptions, the estimates are meaningfully positive."*

They perform no tests, no residual analyses, give no Bayes model selection criteria, no posterior model probabilities, and make no mention of order selection methods. In short, they give us nothing but the summary statement DJW_3. And what a howler DJW_3 is! I leave it to the reader to assess for himself the statistical import of their phrase "meaningfully positive."

Let us now see what a statistical analysis produces. Suppose we employ AIC and BIC order selection criteria in fitting versions of (12) for different values of k. Table 4 gives the value of k selected by these two methods for each of the Nelson-Plosser series.

TABLE 4

Results of Model Order Selection Criteria

Series	Value of lag length k	
	AIC	BIC
Real GNP	1	1
Nominal GNP	1	1
Real GNP per capita	1	1
Industrial Production	1	1
Employment	3	1
Unemployment	3	3
GNP Deflator	1	1
CPI	5	5
Nominal Wages	1	1
Real Wages	1	1
Money Stock	1	1
Velocity	1	1
Bond Yields	2	2
Stock Prices	1	1

It is apparent that standard order selection methods actually *favor* the model (12) with $k = 1$ for many of the series. Indeed, BIC selects $k = 1$ for 11 of the 14 series. Clearly, the choice made in "Critics" was relevant as an objective correlative and is justified by formal selection criteria. DJW^2 are on even shakier grounds here in disputing the choices of model specification that were made in "Critics" than they were in DJW^1 when they preselected $k = 3$ and kept rigidly to that specification.

DJW^2 argue that the presence of serial correlation of the errors in simple AR(1) models tends to bias classical tests in favor of the unit root hypothesis. They cite the development of general unit root tests (such as those in Phillips (1987a) and Phillips-Perron (1988)) as a remedy for this and speak of their AR(3) as "an obvious remedy" in the same spirit. What they neglect to mention is the critical fact that empirical test results with classical methods have shown a remarkable robustness to model specification. By contrast, as "Critics" demonstrates, there is real sensitivity in Bayesian analyses. Classical methods compensate for the increasing downward bias in the MLE $\hat{\rho}$ that occurs as we add more lags because these methods take the sampling distribution of $\hat{\rho}$ into account. Bayesian methods do not compensate in this way and thereby inherit the bias. The Jeffreys-prior approach goes some way in compensating for this but is generally not up to this task in models with many lags and deterministic trends. Hence, there are major problems of bias and fragility to model specification in the routine use of Bayesian methods in this context. "Critics" made this warning loud and clear. DJW choose to ignore it and do so in the face of the evidence.

DJW point to model specifications they have employed in other work as evidence of robustness. In particular, they point to the structural components model representation used in DeJong and Whiteman (1989b, 1991), hereafter $DJW^{3,4}$. Yet in that work they again use a rigid lag specification (in effect, $k = 3$ again) and make no attempt to examine sensitivities to the choice of k. Thus, arguments identical to those made in the preceding paragraph continue to apply. In fact, in recent work with Denis Kwiatkowski and Peter Schmidt (1991) we have found that choice of lag has a major impact on classical tests of stationarity (i.e. $\sigma_\eta^2 = 0$ in DJW^2 notation) in such structural components models. This indeed bodes ill for the Bayesian approach employed in $DJW^{3,4}$.

(ii) *Priors*

DJW's repeated false claims that my description of their prior is incorrect have been rebutted earlier in this response. They make two additional points about the "Critics"-prior. First they evoke a strong subjective sentiment against the shape of the "Critics"-prior.

(DJW_4) *"Phillips' prior assigns relatively enormous weight to explosive roots. The apparent degree of certainty his prior assigns to very rapid exponential growth seems excessive to us."*

DJW tell us that

(DJW_5) *"We certainly are not completely ignorant of ρ or Λ: explosive oscillations (ρ and $\Lambda < -1$) or rapid explosive growth (ρ and $\Lambda > 1.2$, say) do not warrant consideration for real economic time series like the unemployment rate and industrial production. Priors ought to reflect such knowledge; Phillips' prior does not."*

In other words DJW do not wish to profess ignorance. As subjective Bayesians, they may employ priors that reflect whatever information they believe to be relevant. No doubt it is based on knowledge of typical time series trajectories for these economic data (including, of course, the inevitably well known information about the behavior of these US series over the given period!). But what of the trajectories that might have been? Explosive and explosive oscillatory behavior are not totally excluded possibilities especially over subperiods of data. If such behavior were to occur we know *a priori* that the data would be much more informative. Hence, the form of the "Critics"-prior. Subjunctive reasoning such as this seems not to sit well with DJW.

To my reading, their own position as articulated in DJW_5 is simply a dogmatic assertion. It is curious to find this degree of intolerance in Bayesians who by their very actions are subjective. If one elects to take the subjective Bayesian route one should also accept that others are entitled to their own priors, more especially when they are supported by good reason. Edward Leamer in his comments provides a good example in this regard. DJW are eager to get this principle established when it works in their favor, asserting as they do in their leading footnote of DJW^1

(DJW_6) *"The priors, and any errors, are ours."*

Apparently, as subjective Bayesians, they are happy with their own priors. But not only are they unhappy with my prior, they insist in DJW_5 that *their* knowledge ought to be reflected in it as well. Such double standards inevitably lead to a loss in credibility. To repeat but with a new twist: *de rebus prioribus ad nauseam est disputandum*!

DJW's second objection under this heading is that "Critics" employs approximations to extract analytic posteriors. They are right on this point and I am quite unashamed about it. Approximations are often said to be the soul of science. The approximations used in "Critics" are all well explained with full supporting arguments and references. Little more needs to be said here. But let me remark:

1. The Laplace approximation used in (19) and elsewhere in "Critics" has very good analytic properties -- see, for example, the detailed treatment in Bleistein and Handelsman (1976), a reference given in my earlier work (1983) on this method. The method is also known to work well in practice and I have had past good experience with it both in the (1983) paper and in Holly and Phillips (1979) where the quality of the approximation is numerically evaluated. I have no doubt that it is entirely satisfactory in the present application.

2. The confluent function Ψ in (20) is approximated by the simpler analytic form (22). This form is especially valuable because it produces a direct comparison with the posterior for the simpler AR(1) model. Contrary to the following assertion in DJW

(DJW_7) *"Phillips' approximation is a good one to the extent he is certain that ρ is large,"*

I numerically evaluated the Ψ function representation (20) and found it to be well approximated by (22) except for a small region around $\rho = \hat{\rho}$ where the Ψ function is very difficult to compute accurately. I reported on this matter briefly in "Critics," saying that

($PCBP_1$) *"...computations comparing (20) and (22) show that (22) is quite satisfactory for our present purposes."*

Over and above this, one can show (but I certainly will not attempt it here) that the approximation (22) holds analytically for the very complex case where all of the arguments of the Ψ function are large. In short, I am content to stand by my earlier comment $PCBP_1$.

3. In the multidimensional case the "Critics"-prior is not based on the determinantal form but on the product of the diagonal elements of the information matrix. This is a procedure used by Jeffreys (1961) also and is an acceptable simplification, except in some cases where there are strong interactions between the parameters (as there might be in components models of the type (6) discussed above). In

the case of a model with transient dynamics ($k > 1$) I approximate the diagonal elements corresponding to the transient dynamic coefficients by $1/\sigma^2$. Again, this seems a satisfactory simplification. In both instances the reaction of DJW is to boldly allege

(DJW_8) *"...the cost of this approximation is ignored."*

But "Critics" does recognize that approximations like these have consequences -- see the full paragraph of discussion on this very issue following equation (28). That paragraph ends with the statement

($PCBP_2$) *"An adequate methodology for dealing with this extra degree of complication is now under development and will be reported elsewhere."*

This is hardly tantamount to ignoring an issue! "Critics" is a long paper that requires a sustained effort by the reader. But statements like $PCBP_2$ and the paragraph that precedes it in "Critics" are amongst the easiest to read. I submit that if anyone deserves the allegation leveled in DJW_8 of ignoring issues it is DJW not I.

(iii) *Bias and parameterization*

DJW claim that my arguments about the bias of flat-prior analyses are misdirected as criticisms of DJW^1. They give two reasons.

1. First they insist that their prior on Λ is not flat. Then they argue that because of bias in their decision rule they employed a new "5% prior" to compensate for it. Clearly there is a *nonsequitur* here. First of all there *is* flatness. DJW^1 starts with a flat prior on the autoregressive coefficients. When $k = 1$ (an *empirically* important case as our model selection results show) the prior on Λ is also flat. When $k > 1$, Λ is a nonlinear function of the coefficients and the prior is not flat. But the bias that is revealed in "Critics" is still present. Otherwise DJW would not have had to make recourse to their so called "5% prior." Unfortunately the "5% prior" is totally arbitrary and fails to deal with the underlying problem. It is a prior on the trend coefficient "δ" not ρ, it is based on a small scale (1000 replication) Monte Carlo study with specific values of the parameters and is constructed only to make a coarse adjustment to a tail posterior probability. As shown in "Critics" the bias involves a locational shift that seriously misplaces the entire posterior distribution. For the model with a fitted trend the *expected* posterior probability of $\{\rho \geq 1\}$ is less than 5% -- see equation (24) of "Critics." A much more profound adjustment is necessary to compensate for this bias than DJW's 5% prior. Let me observe, in addition, that discussants who object to the dependence of the "Critics"-prior on the sample size should have a heyday demolishing the "5% prior" -- it depends on specific choices of $T = 50$, $\delta = 0.05$ and $\sigma^2 = 1.0$ in model (6) and relies on Monte Carlo estimates of the single tail probability $P_{DJW}(\rho \geq 0.975)$ from 1000

replications. It is such a victim of specificity that there seems to be little point in giving it further serious attention.

2. DJW claim that the results in "Critics" are biased towards integration because of the parameterization chosen for (12):

(DJW_9) *"Second, it is in fact Phillips' procedure which is biased, in favor of integration. The bias results from the use of ρ, which is a poor approximation to Λ when higher-order AR coefficients are important."*

When $k = 1$, we have $\rho = \Lambda$ and the parameterizations are equivalent, so that the DJW[1] procedure suffers all of the bias problems that are raised in "Critics." When $k > 1$, DJW insist that Λ is the key parameter, that ρ is "a poor approximation to Λ" and that "estimates of ρ are substantially larger than estimates of Λ" (reference Table 2 in DJW[2]).

First let us address the question whether Λ is indeed the key parameter. For this is the premise on which the rest of DJW's assault is based. Since parametric models are all best viewed as approximations, the most sensible way of dealing with this issue in a general way is to look at the behavior of the spectrum (of the series) at the origin, what I have called in earlier work the long-run variance. If $f(\lambda)$ is the spectrum of an integrated series, then its behavior in the neighborhood of the origin is characterized by

$$f(\lambda) \sim c/\lambda^2, \ \lambda \to 0$$

for some constant $c \neq 0$. Now consider the AR(k) model

$$(12)' \qquad y_t = a(L)y_t + \varepsilon_t = \rho y_{t-1} + \Sigma_{j=1}^{k-1}\varphi_j \Delta y_{t-j} + \varepsilon_t, \ \varepsilon_t \equiv \text{iid}(0, \sigma^2).$$

The spectrum of y_t has the following equivalent representations

$$(15) \qquad f(\lambda) = (1/2\pi)\sigma^2\left|1 - \rho e^{i\lambda} - \Sigma_1^{k-1}\varphi_j e^{ij\lambda}(1-e^{i\lambda})\right|^{-2}$$

$$(16) \qquad = (1/2\pi)\sigma^2\left|\Pi_{j=1}^{k}(1 - \lambda_j e^{i\lambda})\right|^{-2},$$

where

$$\lambda(z) = \Pi_{j=1}^{k}(1 - \lambda_j z) = 1 - a(z)$$

and $\Lambda = \max_j|\lambda_j|$. Observe that the long-run variance of y_t is given by

(17) $$\text{lrvar} = 2\pi f(0) = \sigma^2(1-\rho)^{-2} = \sigma^2\left\{\Pi_{j=1}^{k}(1-\lambda_j)\right\}^{-2}.$$

Thus, the key parameter that determines the behavior of the spectrum at the origin is ρ not Λ. In fact, it is not Λ itself but the behavior of all of the roots λ_j , $j = 1, ..., k$ that influence $f(0)$ and the long-run variance of y_t. The premise underlying DJW_9 is therefore a colossal blunder and their criticism of "Critics" on this ground is without foundation. On the contrary, it is the DJW[1] choice of parameterization that is inappropriate.

Let me take the algebra a little further and explore the densities of λ_j induced by flat priors on the autoregressive coefficients. In transforming to latent roots the jacobian is typically a linkage factor of the form $|\prod_{j<i}(\lambda_j - \lambda_i)|$. Let us take the case $k = 2$ with roots $\Lambda = \lambda_1 > \lambda_2$, say. The joint density of the roots is

$$\pi(\lambda_1, \lambda_2) \propto (\lambda_1 - \lambda_2).$$

Suppose we now truncate the distribution of these roots, as in DJW[1], between limits c_* and c^*, i.e.

$$c_* \le \lambda_2 < \lambda_1 < c^*.$$

Then the marginal densities are given by

(18) $$\pi(\lambda_1) \propto \int_{c_*}^{\lambda_1}(\lambda_1-\lambda_2)d\lambda_2 = \lambda_1(\lambda_1-c_*) - (1/2)(\lambda_1^2-c_*^2) = (1/2)(\lambda_1-c_*)^2,$$

and

(19) $$\pi(\lambda_2) \propto \int_{\lambda_2}^{c^*}(\lambda_1-\lambda_2)d\lambda_1 = (1/2)(c^{*2}-\lambda_2^2) - \lambda_2(c^*-\lambda_2) = (1/2)(c^*-\lambda_2)^2.$$

As is apparent from (18) the density of $\Lambda = \lambda_1$ rises like a quadratic over the range from c_* to c^* reaching its maximum at the upper limit c^*. In this respect it is entirely analogous to the DJW[1] prior for Λ (implied, as here, by a *flat* prior on the autoregressions coefficients). Note that an extensive Monte Carlo study is not needed to determine this behavior. By contrast, the density (19) of λ_2 falls like a quadratic over the range from c_* to c^* reaching a minimum at the upper limit.

The upshot of this analysis is straightforward. The so-called informative upward sloping prior on Λ which DJW[1,2] makes such a strong point about is obtained at the cost of a downward sloping prior for the second root λ_2. The second density, in effect, compensates symmetrically for the first and the net effect is flatness, with which the analysis started. In consequence, the suggestion in DJW[1,2] that their prior favors nonstationarity is purely illusory.

Note also that $\rho = \lambda_1 + \lambda_2 - \lambda_1\lambda_2$ and thus $\rho > \lambda_1 = \Lambda$ provided $\lambda_1 < 1$ and $\lambda_2 > 0$. This explains all the outcomes in Table 2 of DJW^2. For all series except bond yields we have $\hat{\rho} > \hat{\Lambda}$. For bond yields we have $\hat{\Lambda} = 1.051 > 1.0$ and then $\hat{\rho} < \hat{\Lambda}$. Again, the results obtained in DJW^2 numerically are the consequence of some simple algebra.

To sum up, DJW_9 and often repeated similar statements in DJW^2 are the consequence of a fundamental blunder concerning the parameterization of long-run behavior. In point of fact, DJW_9 should be turned on its head! The key parameter is ρ, Λ is sometimes a poor approximation to it and the results of DJW^1 are therefore biased even further in favor of stationarity.

(iv) *Inference*

DJW express surprise that the inferences in "Critics" do not differ more substantially from theirs. I suspect this impression is gained by eyeballing the posteriors shown in Figures 4(i)-4(xiv) of "Critics" and by neglecting the posteriors for the AR(1) + trend model (which as I have discussed earlier is unjustified). No doubt they did expect bigger surprises, given the form of the "Critics" priors to which their strong opposition, DJW, has already been recorded (see DJW_4) and discussed.

Notwithstanding their interpretations of the empirical results in "Critics," the differences are substantial in both absolute and relative terms for many of the series, as I pointed out earlier in response to KS. For instance, we have

TABLE 5

Posterior Probabilities from "Critics"

Series	$P_J(\rho \geq 0.975)$	$P_{DJW}(\Lambda > 0.975)$
*Stock Prices	0.278	0.040
*Nominal GNP	0.141	0.020
*Consumer Prices	0.652	0.196
Industrial Production	0.192	0.001
*Nominal Wages	0.100	0.018

For four of these series (asterisked above) the inference of stochastic nonstationarity is strongly confirmed by the Phillips-Ploberger test (see Table 3 earlier in the paper).

DJW find the "Critics" posteriors for the unemployment rate and for industrial production to be "absurd." These posteriors have a strong dominant mode around $\rho = 0.70$, 0.80 and a minor mode around $\rho = 1.25$, 1.15. It is the latter that DJW find absurd. As I said in "Critics"

(PCBP$_3$) *"...objective Bayesian analysis of stochastic trends will sometimes produce outcomes that are quite ambiguous due to a widely dispersed bimodality in the posterior distribution. In these cases, Bayesian methods reproduce in their own way a type of uncertainty that we normally associate with low discriminatory power in classical statistical tests."*

No doubt disjoint Bayes confidence sets are as disconcerting to some, as disjoint classical confidence intervals were disconcerting to Sims (1988). With respect to these two series much of the posterior probability is concentrated about a point (around $\hat{\rho}$) that is well into the stationary region. The density goes virtually to zero as $\rho \rightarrow 1$ and then slowly increases to a low additional mode for $\rho > 1.0$. My interpretation is that the dominant mode attaches a substantial probability to values of ρ considerably below 1 and the second mode indicates that there is some smaller probability that the data could have been generated by a model with a much larger ρ, without being at all specific about its possible value (due to the elongated, platykurtic shape of the second mode). Thus, bimodality of this form is a signal that there is more uncertainty about ρ than the usual F-posterior (which is centered on $\hat{\rho}$ and very close to the primary mode of the "Critics" posterior) would indicate on its own. In other words, the objective analysis is helpful because it signals a possible fragility that would not otherwise be apparent!

This interpretation of the empirical findings of "Critics" is in accord with Berger (1985, p. 125) who states:

> *"... in attempting to achieve objectivity, there is no better way to go than Bayesian analysis with noninformative priors. We will not repeat the arguments here but should mention the other side of the coin – when different reasonable priors yield substantially different answers, can it be right to state that there is a single answer? Would it not be better to admit that there is scientific uncertainty, with the conclusion depending on prior belief?"*

g.3. Termination

"A hit, a very palpable hit!" (William Shakespeare: Hamlet)

DJW chose two epigraphs to lead their comment DJW^2: one by T. S. Eliot from the same source as the header I selected for "Critics"; the other, a well known line by Winston Churchill. Churchill was used to signal their contention that "Critics" was wide of its mark. Far from losing the ground they believed they had won in DJW^1, DJW became exhilarated by a mistaken belief that they had consolidated their victory. To wit: the case for trend stationarity seemed stronger than they had thought.

The engagement in (g.2) should quickly disabuse DJW of these fancies of triumph. As Shakespeare would have it, they have sustained some very palpable hits. DJW leave the war that they initiated suffering the indignity that they have committed misquotation (DJW_1), false imputation (DJW_2) and unjust allegation (DJW_8). Their flat prior methodology has been shown to suffer substantial and irrefutable bias, which is exacerbated by their own poor choice of parameterization. In $DJW^{1,2,3,4}$ they persist in imposing rigid model specifications, they elect to ignore the substantial evidence of sensitivity to lag specification and they resist strong arguments for fragility analyses that would alert readers to the shortcomings in their work. In disputing the value of reporting results from AR(1) specifications, they fly in the face of statistical evidence from model selection criteria and they commit statistical howlers like DJW_3 in an attempt to muster a last ditch defence. In short, their scholarship is questionable, their flat prior methodology stands in ruins and their parameterization is discredited. I can only explain their illusion of Churchillian exhilaration with the thought that the nervous system of their atrophied methodology is now dysfunctional and incapable of registering the shock of a direct hit. In such a state as this ignorance is indeed bliss.

In finale, let us turn to the first epigraph of DJW^2, written by T. S. Eliot. No doubt the purpose of DJW is to take some of the sting out of my own criticism by being seen to strike me publicly with my own stick (*viz.* an Eliot header). Leading off DJW^2, Eliot tells us that when he comments on authors whose work he dislikes, his views are to say the least highly disputable. Throughout DJW^2, DJW project an unmistakable dislike for my work -- the comments that they have turned in are indeed a polemic. The inescapable exegesis is that it is their own views that are highly disputable, not mine! Of course, I wholeheartedly agree. Their epigraph from T. S. Eliot becomes the epitaph by which we are invited by them to commemorate their own misunderstandings. This is surely a malapropism of colossal proportion that sinks the remaining credibility of their polemic like a stone!

(h) Christopher A. Sims (CAS)

Sims (1988), hereafter CAS^1, and Sims-Uhlig (1988/1991), hereafter SU, were strident in their criticism of classical unit root econometrics and unqualified in their claim that Bayesian methods offers a superior alternative. In his comments (hereafter, CAS^2) on "Critics," CAS is less judgmental about classical methods and now openly recognizes weaknesses in Bayesian methodology, especially as it has been applied in the debate over trend versus difference stationarity. This is a shift of ground that I welcome.

Notwithstanding this shift in position, CAS expresses disagreement with many of the explicit suggestions in "Critics" concerning the conduct of Bayesian inference and he reasserts some claims made in CAS^1 and SU concerning the clash between the likelihood principle (LP) and conventional sampling theory asymptotics. Taking an overview of CAS^2, I believe one major reason for our continuing disagreement is now quite clear.

Like many of the earlier discussants, Sims' point of departure is subjective. Sims' priors are the embodiment of his personal beliefs and his disagreement with the "Critics"-priors stems from this perspective. On this point of dissension my reply is a simple one. Subjective Bayesian analysis is an admissible mode of personal statistical inference but if the results of such analysis are to be considered seriously by others, efforts must be made to calibrate the results against some objective correlative. Readers of scientific research are entitled to this type of sensitivity analysis. If it is not undertaken, the results of a Bayesian analysis are, like the prior, entirely subjective and must be treated as such. They are then of much less relevance to others and their scientific value is correspondingly diminished. "Critics" showed that in time series models Bayesian posteriors have an endemic bias that is inherited from the MLE. Subjective analyses of the type considered in $CAS^{1,2}$ and SU make no effort to compensate for this bias and should therefore be treated with even more caution than usual.

The second source of our continuing disagreement concerns the use of the LP in inference with time series. CAS is content to apply the LP in time series models with no attention to any of the implications of living on a particular time series trajectory. In doing so, he conforms with traditional Bayesian thinking and inference proceeds in a mechanical way by the application of Bayes theorem. In "Critics" I questioned the mechanical use of this apparatus in a time series setting and pointed out that conditioning on sample moments that carry information about the parameters is not innocuous. Phillips-Ploberger (1991) fully explores the effect of this conditioning and shows that Bayes inference along these traditional lines implies the existence of a model, what we call the Bayes model, that is very different from the classical model with which the analysis started. This change of reference frame must be taken into account in comparing Bayes and classical methods. $CAS^{1,2}$ and SU do not, of course, do so and on this point our disagreement is a fundamental one. Since the technical arguments in Phillips-Ploberger are rigorous, there is no doubt which is the correct position on this point. I shall take the issue up again below.

Like earlier discussants, CAS gives attention to only part of the "Critics" paper in his comments, notably items II(i), II(iii) and II(iv). The other items are either ignored or addressed only tangentially. I find the omission of II(v) and II(vi) most surprising. II(v) is surely rather uncontroversial -- the Laplace approximation is a useful device and has many potential applications in Bayesian econometrics. II(vi) is also uncontroversial. The sampling properties of Bayes procedures are of interest and "Critics" documents the poor sampling performance of flat prior Bayesian methods in time series models with a unit root. We need to take such findings into account when we assess the merit of different methods. I am puzzled why CAS^2 has nothing to say on this evidence, yet still argues the advantages of a flat prior methodology.

To organize the remainder of this reply to CAS, I shall respond to the points raised in the order in which they appear in CAS^2.

1. CAS^2 reasserts the point made earlier in SU that

(CAS_1) *"... reporting of statistical results ought to be conceived of as communicating information about the shape of the likelihood."*

This is a very old perspective on inference and one that was forcefully argued in the Barnard, Jenkins and Winsten (1962) paper (BJW) that I referenced and discussed earlier in response to KS in II(a). BJW is uncited in $CAS^{1,2}$ and SU, yet I believe it is the original source of this approach to inference in time series. BJW tell us:

(BJW_1) *"... when, on the evidence of a given set of data, the plausibilities of various hypotheses are to be compared, then the primary inference is provided by the likelihood function. The suggestion is made that the practice of looking at the whole course of the likelihood function should receive much more attention."*

Bayesians, of course, automatically follow this advice under flat priors. CAS certainly follows the maxim when he tells us in CAS^2 that

(CAS_2) *"A scientific-reporting perspective suggests that the aim of statistical research reports is to summarize the likelihood. Since a flat-prior posterior is just the likelihood normalized to sum to one, it has direct appeal from this perspective."*

In this sense there is nothing new in $CAS^{1,2}$ and SU about likelihood-based foundations of inference.

I think it fair to say that the latter suggestion in BJW_1, at least to the extent that it is practically feasible, is now also considered to be good practice by classical researchers. The essential difference between the Bayesian and classical approach is that classical hypothesis testing theory does not accept the first maxim of BJW_1, *viz.* that the likelihood function itself is the basis of assessing the plausibility of different hypotheses. Classical theory recognizes the importance of trajectories that might have been and thereby gives weight to such possibilities in the future. Classical theory recognizes weaknesses in the sampling properties of the MLE and suggests procedures that compensate for these. Flat-prior Bayesian methods ignore these issues and this is the major reason why the BJW prescriptions found little support amongst that paper's many discussants at the time. It is notable that BJW's prescriptions for time series have received few followers in statistics. Their recrudescence in recent econometrics work is, therefore, something of a curiosity, more especially since BJW and its attendant debate have been left unmentioned!

2. CAS^2 repeats the argument in SU that there is no prior that will rationalize classical p-values as posterior probabilities for all sample trajectories and then makes the often-repeated claim:

(CAS_3) *This means that treating p-values as probabilities is incoherent."*

It is a very old argument that is put forward in support of the Bayesian approach that to be rational or coherent (with underlying axioms of individual preference) any statistical analysis must correspond to a Bayesian analysis. Readers will find a detailed discussion of the matter in Berger's (1985, Ch. 4) text and a recent treatment and discussion in Berger and Sellke (1987). I do not see that $CAS^{1,2}$ or SU add anything to the preexisting debate on this topic and I find their position unpersuasive and somewhat contradictory given the advice of Sims (1989) concerning suitable critical values in unit root tests.

Note first that $CAS^{1,2}$ and SU all omit reference to important recent work that shows it is possible to reconcile evidence from Bayesian and classical approaches in one sided testing problems. Cosella and Berger (1987), for instance, show that for classes of impartial priors in one sided location tests there is equality between the infimum of Bayes posterior probabilities that H_0 is true and the classical p-value. Thus, for a certain class of problems the classical p-value is on the boundary of the Bayesian posterior evidence. Some related arguments were put forward earlier by DeGroot (1973). This work contradicts the thrust of the argument behind CAS_3 because it provides a very clear sense in which Bayesian and classical measures of evidence can be reconciled. The one sided tests that are used in assessing evidence for the presence of a unit root against stationarity make this theory of reconciliation very relevant to unit root tests. I personally find this theory much more persuasive than the simulation exercises conducted in SU.

Notwithstanding the above arguments, there is of course no reason to insist on a Bayesian interpretation of a classical p-value. In classical theory the concept of a probability that the null hypothesis is true has no meaning, whereas the interpretation of the p-value as a sampling error rate is objective and well understood.

One further point on this matter should be made. Bayesians often argue that p-values are misleading measures of evidence provided by the data against the null hypothesis. In particular, it is often argued that p-values *overstate* the evidence against the null hypothesis. Berger and Sellke (1987) give a recent discussion of the matter and quote Jeffreys (1980) as saying

(J_1) *"... that differences up to twice the standard error usually disappear when more or better observations become available and that those of three or more times usually persist."*

In other words, p-values of around $p = 0.05$ for a t-ratio of 2.0 overstate the evidence against the null and that in J_1's assessment a t-ratio of 3.0 is more realistic for the purposes of rejection of the null. Note that as far as tests of a unit root are concerned, classical unit root asymptotic theory puts the t-ratios required for rejection around 3.0, as distinct from classical stationary asymptotics which put the value at the more usual 2.0. In this context, Sims (1989) recommendation to economists to use critical values around 2.0 in unit root tests seems curiously at odds with his views about the irrationality of p-values. Given the evidence provided by Berger and Sellke (1987) and others on the misleading nature of p-values and the recommendation of Jeffreys in J_1, it would be natural to expect Bayesians to be happier with critical values of 3.0 rather than 2.0.

Finally, on this matter of the rationality of Bayes methods let me quote Berger (1985, p. 121) who tells us that

> *"A Bayesian analysis may be 'rational' in the weak axiomatic sense, yet be terrible in a practical sense if an inappropriate prior distribution is used."*

In short, even Bayesians themselves are sensitive to the inferential distortions that can ensue from the use of poor priors. The blessing of rationality seems to be of little practical advantage when the maxim for the practitioner is: *De rebus prioribus caveat emptor!*

3. CAS^2 repeats the claim made in SU that "classical methods in a sense ignore information" about $\hat{\sigma}_e = \hat{\sigma}(\Sigma_1^n y_{t-1}^2)^{-1/2}$. I rebutted that claim in Section 2(d) of "Critics," pointing out, *inter alia*, that certain classical tests actually depend on $\hat{\sigma}_e^2$ and are known to be most powerful in a neighborhood of the alternative. I see no argument in CAS^2 against this point.

CAS^2 does confront the argument made in "Critics" that Bayesian conditioning is not innocuous and CAS^2 also suggests that my arguments about the effects of conditioning on the sample moment $A_n = \Sigma_1^n y_{t-1}^2$ lead to a "semantic confusion." CAS^2 tells us that

(CAS_4) *"Uhlig and I tried to provide some intuition for how it can be that, despite the downward bias in the OLS estimate $\hat{\rho}$ of ρ, the likelihood for ρ turns out to be symmetric about $\hat{\rho}$."*

The real reason for the symmetric Gaussian shape of the likelihood for ρ (and hence that of the flat prior posterior considered in SU) is that the density from which this likelihood is obtained is taken with respect to a different probability measure, one that arises specifically because of data conditioning. The argument is laid out in full in the Phillips-Ploberger (1991) paper cited earlier. But I shall bring attention to the central issues here. For the AR(1) model as in equation (1) of "Critics," let $h = \rho - 1$, P_n^ρ be the probability measure of $Y_n = \{y_t\}_1^n$ and $P_n = P_n^1$. Then, the likelihood function given Y_n can be written in terms of h as

$$L_n = dP_n^\rho / dP_n = \exp\left\{(1/2)\hat{h}_n^2 A_n\right\}\exp\left\{-(1/2)(\rho - \hat{\rho})^2 A_n\right\}, \qquad (20)$$

(see equation (39) of Phillips-Ploberger (1991)) where $h_n = \hat{\rho}_n - 1$ and $\hat{\rho}_n$ is the OLS/MLE of ρ based on Y_n. From (20) it is apparent that only the second factor for L_n is important for likelihood-based inference about ρ and this is the factor that produces the symmetric Gaussian shape about $\hat{\rho}_n$. Note that L_n may be written in a more revealing manner as follows:

$$L_n = \left[A_n^{-1/2}\exp\left\{(1/2)\hat{h}_n^2 A_n\right\}\right]N(\hat{\rho}_n, A_n^{-1}) \tag{21}$$

$$\propto N(\hat{\rho}_n, A_n^{-1}). \tag{22}$$

In (22) we get the direct Gaussian posterior for ρ about $\hat{\rho}_n$ that applies under a flat prior. Note from (20) and (21) that deviations of ρ from $\hat{\rho}_n$ are measured in units of $A_n = \Sigma_1^n y_{t-1}^2$. This changes the geometry of inference. Conditional on A_n the likelihood for ρ and the posterior is Gaussian, but the conditioning under which this is true is certainly not innocuous. The passage to the posterior via the proportionality sign that appears innocuously in (22) eliminates the first factor of (21). This factor, as the Phillips-Ploberger paper shows, changes the reference measure from P_n (the measure of the unit root model) to a conditional Bayes model measure in which $\hat{\rho}_n$ figures prominently, *viz.*

$$y_{n+1} = \hat{\rho}_n y_n + u_n. \tag{23}$$

There is no "semantic confusion" here. Bayes methods involve implicitly a change in the model and a change in the underlying probability measure. This is certainly not innocuous and, as Phillips-Ploberger show, has major implications for inference.

When CAS says

(CAS_5) *"I assume Phillips accepts Bayes rule as a formula for calculating conditional distributions from a given joint distribution – this formula is not what is at issue between Bayesians and frequentists"*

my response is this: Yes I accept the formula but I also accept what the operation of Bayes rule involves in terms of the implied model and the reference measure. One major import of the Phillips-Ploberger analysis is to make these implications of Bayes rule in time series models explicit, it would seem, for the first time.

Finally, the artificial experiment described in SU where values of ρ are drawn at random from a uniform distribution is, in my view, irrelevant for conditional inference. If ρ is truly random and its distribution known to be uniform then data like Y_n can tell us no more about ρ. By contrast when nothing is known about ρ, Bayes inference is entirely trajectory based on Y_n. The Bayes model that is implied by the operation of Bayes rule is then evolutive and has its own probability measure. In other words, when the true mechanism that generates ρ is unknown, the Bayes approach is to continually revise the model as the trajectory evolves, just as in (23) above. This is the reason why the conditional distribution of ρ in Bayes inference is symmetric about $\hat{\rho}_n$ and this is why conditional inference is fundamentally different in the probabilistic sense. From my reading of $CAS^{1,2}$ and SU I see no appreciation of this point at all. On this matter, then, we remain fundamentally divided.

4. In "Critics" I pointed out what I saw as the many disadvantages of using levels VAR's in empirical research. CAS says he is unconvinced by these, while accepting that the points made have some validity. He concludes that my own suggestions amount

(CAS_6) *"... to a recommendation in favor of rose-colored glasses."*

These issues over VAR's are tangential to our main methodological debate over the specifics of good Bayesian inference. I stand by the statements I made in "Critics" and shall therefore confine myself in this reply to a few additional points.

(i) VAR's in levels are known to perform poorly even in forecasting exercises unless they are heavily restricted by prior information that includes unit root priors. Such prior information is, in my perspective, at least as arbitrary as the identifying restrictions in structural models claimed earlier in Sims (1980) to be incredible. Analysis of impulse responses requires further identifying information. Moreover, implementation of VAR priors is usually very mechanical (e.g. the first order autoregressive coefficient is typically pulled towards unity for all variables in the model) and therefore almost certainly wrong, ignoring for instance the presence of some degree of cointegration in most macroeconomic data. Thus, levels VAR's inevitably pay a penalty for overparameterization (in terms of additional variability) and another penalty for misspecification (insistence on an excessive number of prior unit roots). Finally, when VAR's do implicitly estimate long run structural relationships they do so with bias and that bias is known to be nonnegligible in finite samples.

(ii) Levels VAR's provide an unsatisfactory basis for classical causality tests. Sims, Stock and Watson (1990) originally pointed out some of the difficulties here and a recent paper, Toda and Phillips (1991), provides a complete study of causal inference in this context, ending up with the conclusion that causality testing in levels VAR's is not to be recommended.

(iii) ECM's do offer consistent testing procedures for unit roots, the dimension of the cointegration space and optimal estimation of the cointegrating vectors. These systems also commence from an atheoretical unrestricted format that is actually identical up to parameterization to a levels VAR. In the face of this latter correspondence, the decision to work with levels VAR's and their induced impulse responses is, in effect, a decision against statistical inference. If to use ECM's and more parsimoniously parameterized structural models is to be regarded as a recommendation in favor of rose-colored glasses, as stated in CAS_6, then I much prefer this to the alternative that CAS seems to embrace of ignoring statistical inference issues altogether. The half way house of Bayesian inference in multivariable time series models would be a desirable meeting place between these two alternatives.

(iv) Structural ECM models are now known to be capable of parsimoniously encompassing VAR's, as recent work of Hendry and Mizon (1990) shows. In such cases the arguments in favor of using structural ECM models in empirical work rather than unrestricted VAR's seem to me to be compelling.

5. CAS^2 is critical of the "Critics"-prior ($\pi^T(\rho)$ in the notation of II(d) above) telling us that

(CAS_7) *"From a likelihood-reporting perspective the sample-size dependence of these priors is perverse."*

and

(CAS_8) *"The heavy weight placed by the prior on extremely explosive models is also unreasonable."*

This opposition is grounded on subjectivist Bayesian principles, whereby the prior is deemed to represent the prior beliefs of the investigator. I have dealt with similar subjectivist comments from other discussants earlier (see, especially, my comments on KS, EL and DJP). The need for objective Bayesian inference in time series to which "Critics" responded is given little attention in CAS^2.

CAS is especially critical of the sample size dependence of $\pi^T(\rho)$, telling us that

(CAS_9) *"... since no reader will have a prior that varies with the sample size, it must be rare that the Jeffreys prior is close to a reader's beliefs."*

But the prior $\pi^T(\rho)$ is not intended to represent a reader's belief in any conventional sense and it is certainly not subjective. So the attempt in CAS^2 to force the approach within subjectivist Bayesian thinking is certain to fail. Note that I made no attempt to do this in "Critics," but instead argued strongly in favor of its impartiality as a model-based reference prior that provides an objective correlative for other priors like the flat prior.

Why does $\pi^T(\rho)$ depend on T? CAS, KS, DJP and others are genuinely puzzled by this. The answer is simple. *A priori* we know that the information content of autocorrelated data depends on the number of observations and does so in the special way that this information increases nonlinearly with the value of ρ. The sample size dependence of $\pi^T(\rho)$ reflects precisely this idea. In models with independent observations information about a parameter collects uniformly as $T \to \infty$ and objective priors are independent of T. In the AR(1), however, we know that the rate of accumulation of information is explosive in T for $\rho > 1$. This knowledge is incorporated in the "Critics" prior and it compensates for the fact that the unscaled likelihood converges at the same rate in T in local neighborhoods of $\hat{\rho}$. Thus, from (21) we have the following unscaled likelihood for $T = n$

(24) $$\bar{L}_n = \exp\left\{-(1/2)(\rho - \hat{\rho}_n)^2 A_n\right\}.$$

Now let

(25) $$\rho = \hat{\rho}_n + h/a_n, \text{ with } a_n = \pi^n(\rho_0)$$

where ρ_0 is the true value of ρ. For ρ in the neighborhood (25) we have the following (weak) limit on L_n

(26) $$\bar{L} = \exp\left\{-(1/2)h^2 K_0\right\}$$

where K_0 is the (weak) limit of $a_n^{-2}A_n$. Equation (2) is a special case of this for $\rho_0 = 1$ and with $K_0 = \int_0^1 W^2$. Under a flat prior $\bar{L}_n$ delivers the posterior directly with a scaling that is uniform in ρ. Under the "Critics" prior the likelihood is rescaled as $\pi^n(\rho)\bar{L}_n$. This scaling compensates for the symmetry of $\bar{L}_n$ about $\hat{\rho}_n$ and incorporates the knowledge that as n increases information about ρ accumulates much more rapidly on the $\rho > \hat{\rho}_n$ half line than it does for $\rho < \hat{\rho}_n$. The limit, however, is still proportional to $\bar{L}$ because the information in the data (or likelihood) dominates when $n \to \infty$.

The above argument shows that CAS^2 is wrong in claiming that the likelihood does not dominate as $T \to \infty$. Thus, CAS^2 tells us that

(CAS_{10}) *"Whatever the subjective prior, so long as it is characterized by a continuous pdf, the posterior will under usual regularity conditions come to resemble the likelihood function (the flat prior posterior pdf). This argument applies to Jeffreys priors in usual applications as well, since in the usual context Jeffreys priors do not depend on sample size. The Jeffreys priors for this time series application, however, depend so strongly on sample size that this justification for them fails – posteriors from these Jeffreys priors have infinite variance in every sample size, whereas posteriors from any proper subjective prior with continuous and bounded pdf eventually have finite variance that shrinks to zero with sample size."*

But finite variance is not a necessary condition for convergence! It is a common phenomena for classical estimators (e.g. FIML) to have no integer moments in any finite sample size but still converge. The fact that the "Critics" posterior has Cauchy like tails for every finite n does not destroy its convergence properties.

6. I agree with CAS^2 that issues of time aggregation are important. The simplest and most coherent way to deal with such issues is to embed the model in a continuous time framework. Then the sampling interval becomes a parameter in the likelihood and prior, just as the sample size. I raised this possibility earlier in responding to KS and will say no more about it here. Note that if one really does want to use a subjective prior that drops away sharply for $\rho > 1$ the ε-priors given in (3) give a nice way of doing this. Zivot-Phillips (1991) have a very extensive analysis that utilize this type of prior.

7. In models with trend terms CAS argues that

(CA_{11}) *"... the Jeffreys priors retain their same general shape while the classical distribution theory shifts dramatically. When μ and β in Phillips' model with trend are non zero, classical asymptotic theory yields Gaussian distributions for the estimated ρ even in the presence of a unit root."*

The critical hypothesis in CAS_{11} is that $\beta \neq 0$. When $\beta \neq 0$ in a model with $\rho = 1$ there is a quadratic trend in the model, which is quite different from the linear trend that applies for $\rho < 1$. In considering the possibility of a model with deterministic trend and a unit root, it is now conventional to set $\beta = 0$ but retain the trend term in the regression. This is then equivalent to taking a deterministic trend out of the data prior to the analysis of the presence of a unit root. In that case the usual unit root asymptotics apply but with detrended Brownian motions.

An alternative is to employ the more parsimonious components representation (6) in which case we again get nonstandard limit theory, as shown in Schmidt-Phillips (1989). Bayesian theory runs into some difficulty when we employ the components representation because of the degeneracy in the explicit parameterization (7) at $\rho = 1$. The singularity that is induced in the flat prior at $\rho = 1$ by this parameterization is mentioned in CAS^2 and I have discussed it earlier in II(e) in response to SVD. This singularity takes the form $|1-\rho|^{-j}$ ($j = 1, 2$) and produces a nonintegrable improper prior and nonintegrable improper posterior. This anomaly is the consequence of forcing an overparameterized reduced form to accept a degenerate parameterization. The approach I adopted in "Critics" avoids these difficulties. The solution to these difficulties suggested by CAS^2 and SVD is the same -- a proper normal prior on the intercept. This is clever but arbitrary, and sensitivity of the posteriors to this choice is warranted. Figure 3 of CAS^2 shows the marginal posteriors for ρ for two choices of this arbitrary prior against the posteriors for a flat prior and for stationary Jeffreys prior. The graphs demonstrate a striking fragility to the choice of prior. Although they are not given, the tail posterior probabilities for near nonstationarity, i.e. $P(\rho \geq 0.975)$, clearly differ enormously for the different priors, with those from the flat prior being the smallest. I take this as confirmation of the conclusion in "Critics" concerning fragility and an endorsement of the theme of this reply concerning the need for an objective correlative in assessing fragile posteriors.

8. CAS^2 concludes that

(CAS_{12}) *"The argument that flat priors are unreasonable when the model contains constant and trend is stranger than I had realized,"*

but asserts that

(CAS_{13}) *"In the homogeneous model, with no constant or trend term, a flat prior seems sensible."*

Apparently, we still differ on CAS_{13}. There is no doubt that the simulation evidence in "Critics" shows up the poor sampling properties of flat prior Bayesian methods in the homogeneous model as well as the model with deterministic trend. Some procedure that compensates for the bias in the MLE $\hat{\rho}$ is required. The "Critics" prior seems to me to be a good objective mechanism for doing this and, even in a subjective approach, there would seem to be no real argument for not using it as an objective correlative in assessing the fragility of posteriors that are based purely on subjective priors. I see no argument in CAS^2 against this proposal.

To sum up, I think there is good reason to believe that this debate has narrowed the ground that divides us on some aspects of Bayesian inference. CAS has offered his comments on "Critics" in a productive spirit that is characteristically thoughtful and I welcome his further contribution. Much of the ground that still separates our thinking concerns the foundations of Bayesian inference and the operational implications of the likelihood principle and Bayes rule. I hope that we will have a further opportunity to address these differences once the ideas and methods of the Phillips-Ploberger (1991) paper have been assimilated.

III. CONCLUSION

There can be no dispute that scientific interest in Bayesian methods has grown sharply in recent years. The phenomena has affected many disciplines besides econometrics within the larger statistics community. And Bayesian methods of modeling learning mechanisms have begun to attract interest amongst economic theorists as well. The challenge to the applied econometrician is to find good methods of implementing the Bayesian approach in practice and objectively assessing the findings so that they are useful to a wider scientific audience. Time series models and data present special difficulties in this regard. The "Critics" paper pointed out these problems, raised some new perspectives and offered some solutions. The comments of the discussants have added further thoughts and evidence on these issues. I thank them all for their comments and I thank the Editor and the *Journal of Applied Econometrics* for supporting a productive exchange of ideas on this important topic.

IV. REFERENCES

Barnard, G. A., G. M. Jenkins, and C. B. Winsten (1962). "Likelihood inference and time series," *Journal of the Royal Statistical Society*, A, 125, 321-372.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Verlag (2nd Edition).

Berger, J. O. and M. Delampady (1987). "Testing precise hypotheses," *Statistical Science*, 2, 317-352.

Berger, J. O. and T. Sellke (1987). "Testing a point null hypothesis: The irreconcilability of p-values and evidence," *Journal of the American Statistical Association*, 82, 112-122.

Bleistein, N. and R. A. Handelsman (1976). *Asymptotic Expansions of Integrals*. New York: Holt, Rinehart and Winston.

Cosella, G. and R. L. Berger (1987). "Reconciling Bayesian and frequentist evidence in the one-sided testing problem," *Journal of the American Statistical Association*, 82, 106-111.

DeGroot, M. H. (1973). "Doing what comes naturally: Interpreting a tail area as a posterior probability or likelihood ratio," *Journal of the American Statistical Association*, 68, 966-969.

deJong, D. N. and C. H. Whiteman (1989a). "Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle," Working Paper No. 89-4, Department of Economics, University of Iowa. To appear in *Journal of Monetary Economics*.

_______ (1989b). "Trends and cycles as unobserved components in US real GNP: A Bayesian perspective," *Proceedings of the American Statistical Association, Business and Economics Section*, 63-70.

_______ (1991). "The temporal stability of dividends and stock prices: Evidence from the likelihood function," Working Paper No. 89-3, Department of Economics, University of Iowa, to appear in *American Economic Review*.

Ghosh, J. K. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays by Dr. D. Basu*. New York: Springer Verlag.

Hendry, D. F. and G. Mizon (1990). "Evaluating dynamic econometric models by encompassing the VAR," mimeographed, Oxford University.

Hinkley, D. V. (1983). "Can frequentist inferences be very wrong: A conditional 'Yes'," in G. E. P. Box, T. Leonard and C.-F. Wu (eds.), *Statistical Inference, Data Analysis and Robustness*. New York: Academic Press.

Holly, A. and P. C. B. Phillips (1979). "A saddlepoint approximation to the distribution of the k-class estimator in a coefficient in a simultaneous system," *Econometrica*, 47, 1527-1547.

Jeffreys, H. (1961). *Theory of Probability*, 3rd Edition. London: Oxford University Press.

_______ (1980). "Some general points in probability theory," in A. Zellner (ed.), *Bayes Analysis in Econometrics and Statistics*. Amsterdam: North Holland, pp. 451-454.

Kwiatkowski, D., P. C. B. Phillips and P. Schmidt (1990). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root," Cowles Foundation Discussion Paper No. 979.

Lindley, D. V. (1990). "The 1988 Wald Memorial Lectures: The present position in Bayesian statistics," *Statistical Science*, 5, 44-89.

Nelson, C. R. and C. Plosser (1982). "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.

Phillips, P. C. B. (1983). "Marginal densities of instrumental variables estimators in the general single equation case," *Advances in Econometrics*, 2, 1-24.

_______ (1987a). "Time series regression with a unit root," *Econometrica*, 55, 277-301.

_______ (1987b). "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535-547.

_______ (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*.

Phillips, P. C. B. and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.

Phillips, P. C. B. and W. Ploberger (1991). "Time series modeling with a Bayesian frame of reference: I. Concepts and illustrations," Cowles Foundation Discussion Paper No. 980.

_______ (1991b). "Time series modeling with a Bayesian frame of reference: II. General theory and applications," in preparation.

Schwert, G. W. (1989). "Tests for unit roots: A Monte Carlo investigation," *Journal of Business and Economic Statistics*, 7, 147-160.

Schmidt, P. and P. C. B. Phillips (1989). "Testing for a unit root in the presence of deterministic trends," Cowles Foundation Discussion Paper No. 933.

Schotman, P. and H. K. van Dijk (1991). "A Bayesian analysis of the unit root in real exchange rates," *Journal of Econometrics* (forthcoming).

Sims, C. A. (1980). "Macroeconomics and reality," *Econometrica*, 48, 1-48.

_______ (1989). "Comment on S. Durlauf, 'Output persistence, economic structure and the choice of stabilization policy'," *Brookings Papers on Economic Activity*, 1989:2, 125-129.

Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrica*, 58, 113-144.

Sims, C. A. and H. Uhlig (1988/1991). "Understanding unit rooters: A helicopter tour," Federal Reserve Bank of Minneapolis Institute for Empirical Macroeconomics, Discussion Paper No. 4. To appear in *Econometrica*.

Stock, J. H. (1990). "Confidence intervals for the largest autoregressive root in US macroeconomic time series," UC Berkeley manuscript.

Toda, H. and P. C. B. Phillips (1991). "Vector autoregressions and causality," Cowles Foundation Discussion Paper No. 977.

Zivot, E. and P. C. B. Phillips (1991). "A Bayesian analysis of trend determination in economic time series," manuscript, Yale University.

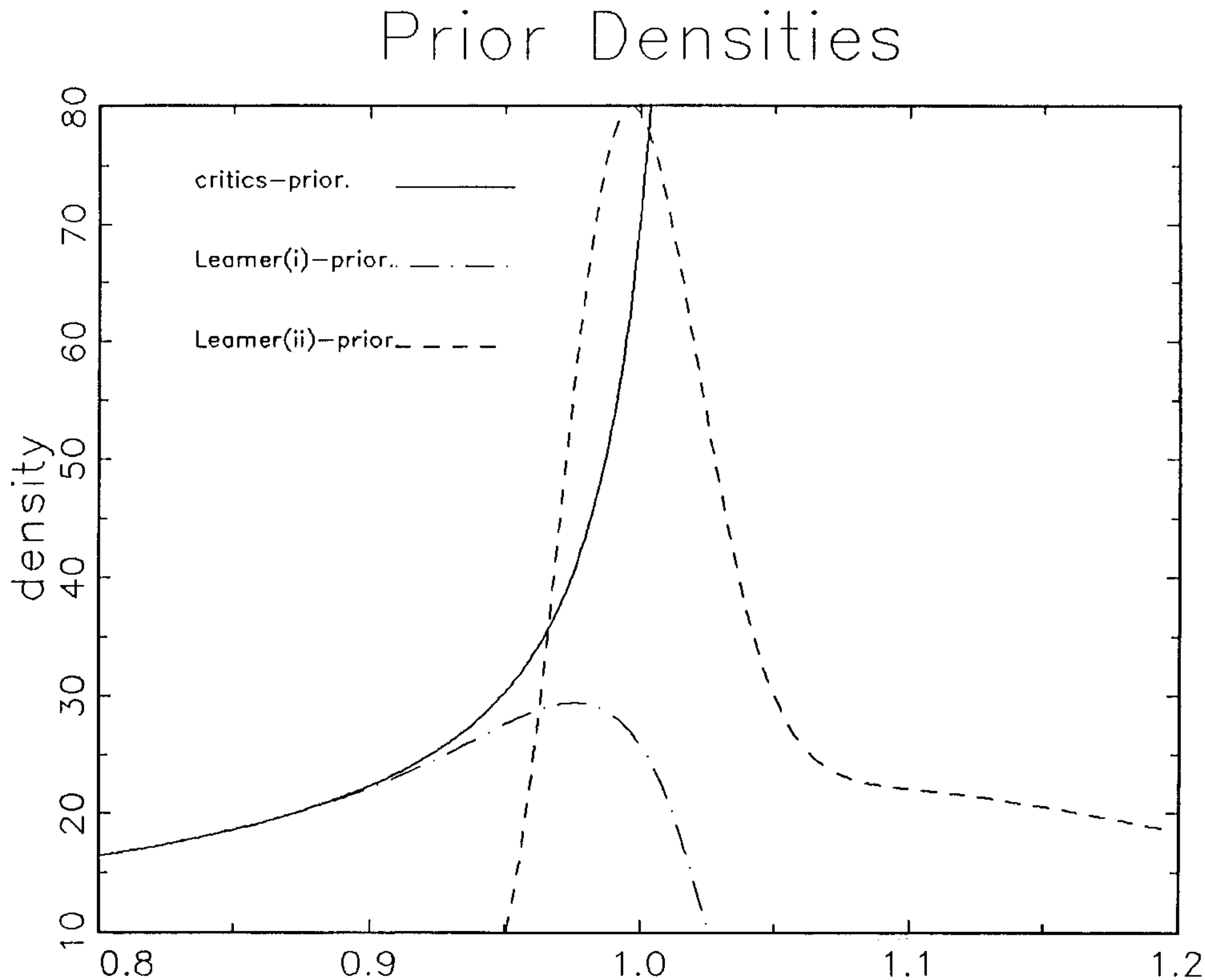

Figure 0(i): Prior Densities

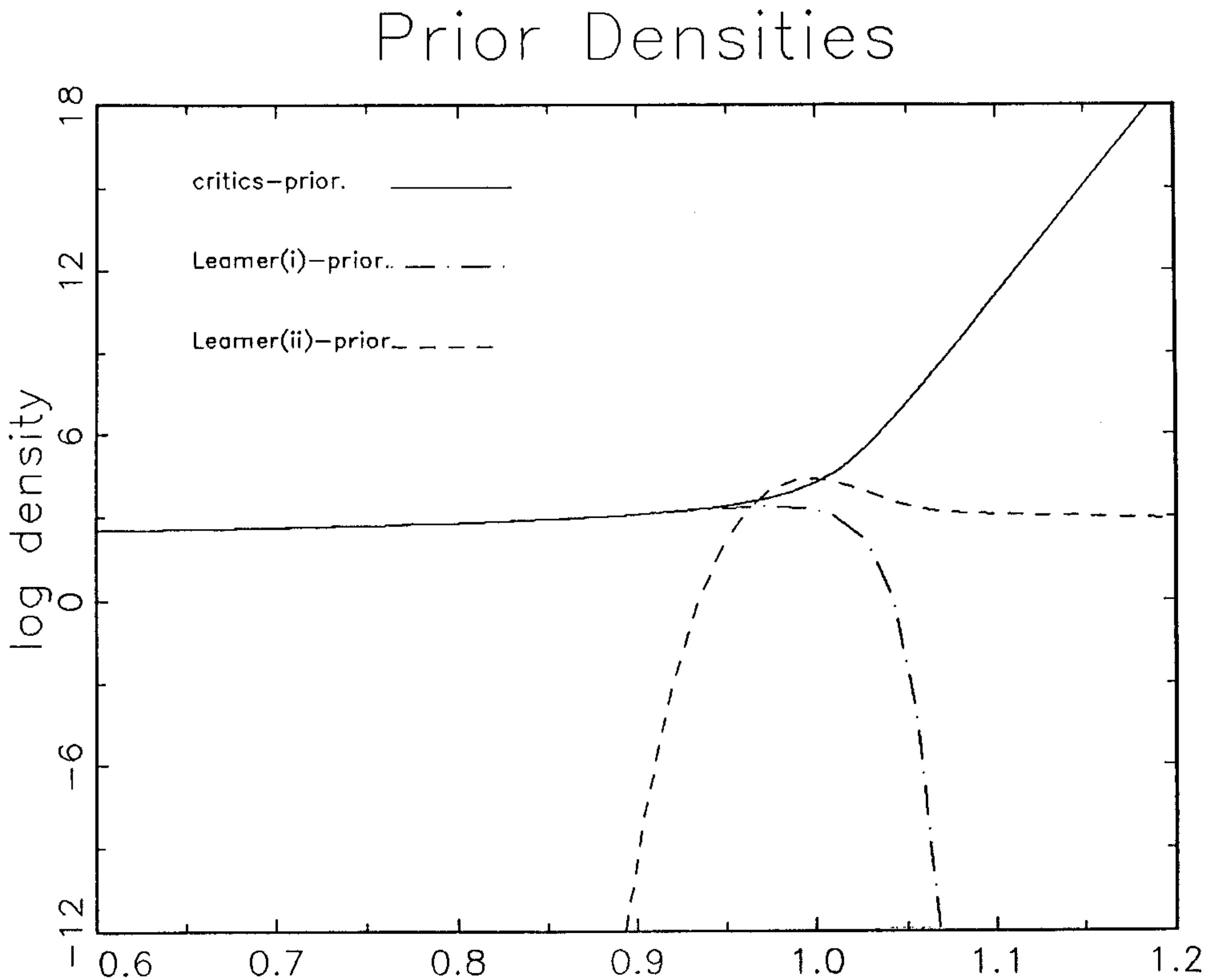

Figure 0(ii): Prior Densities

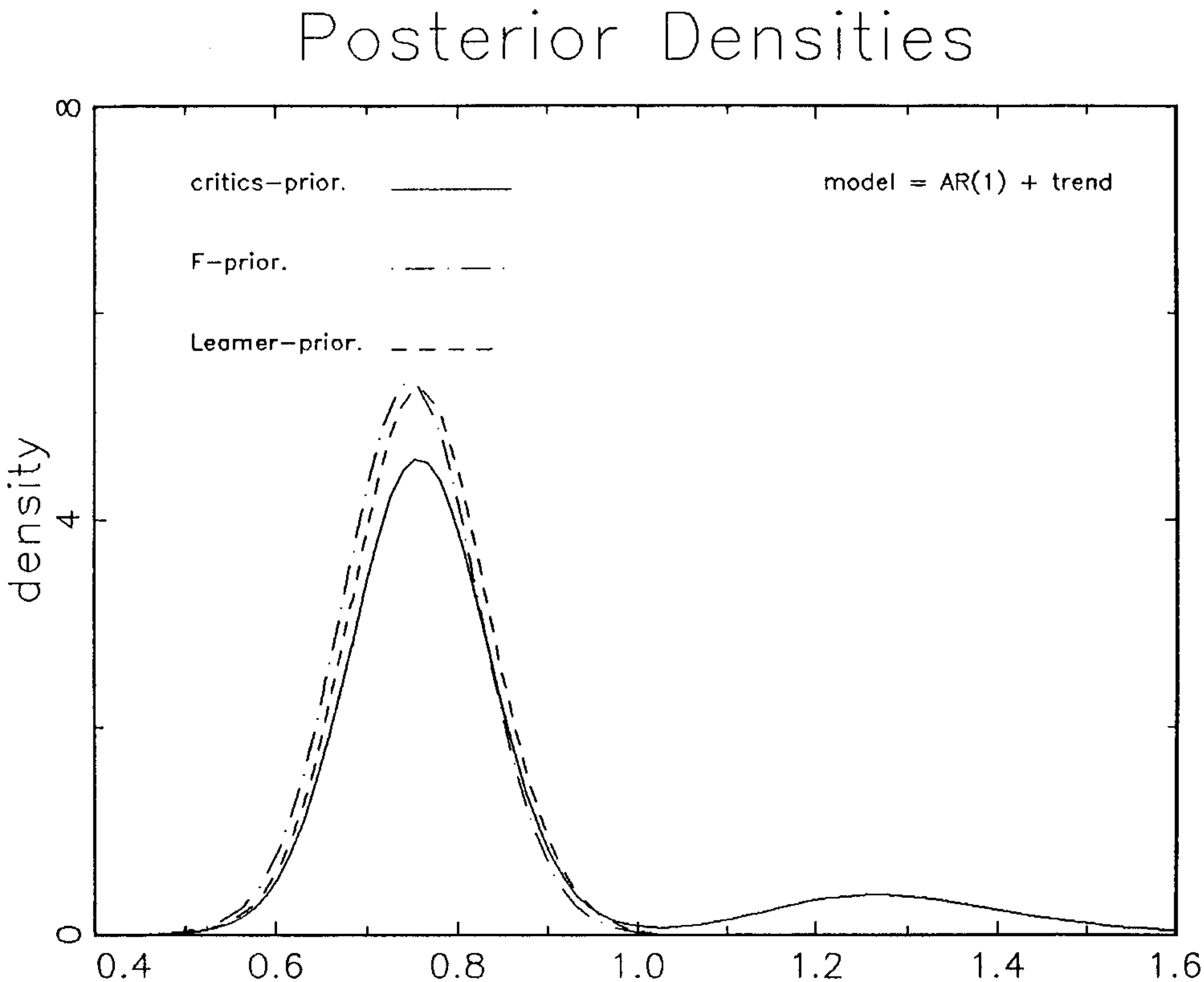

Figure 1: Unemployment rate: 1890-1970

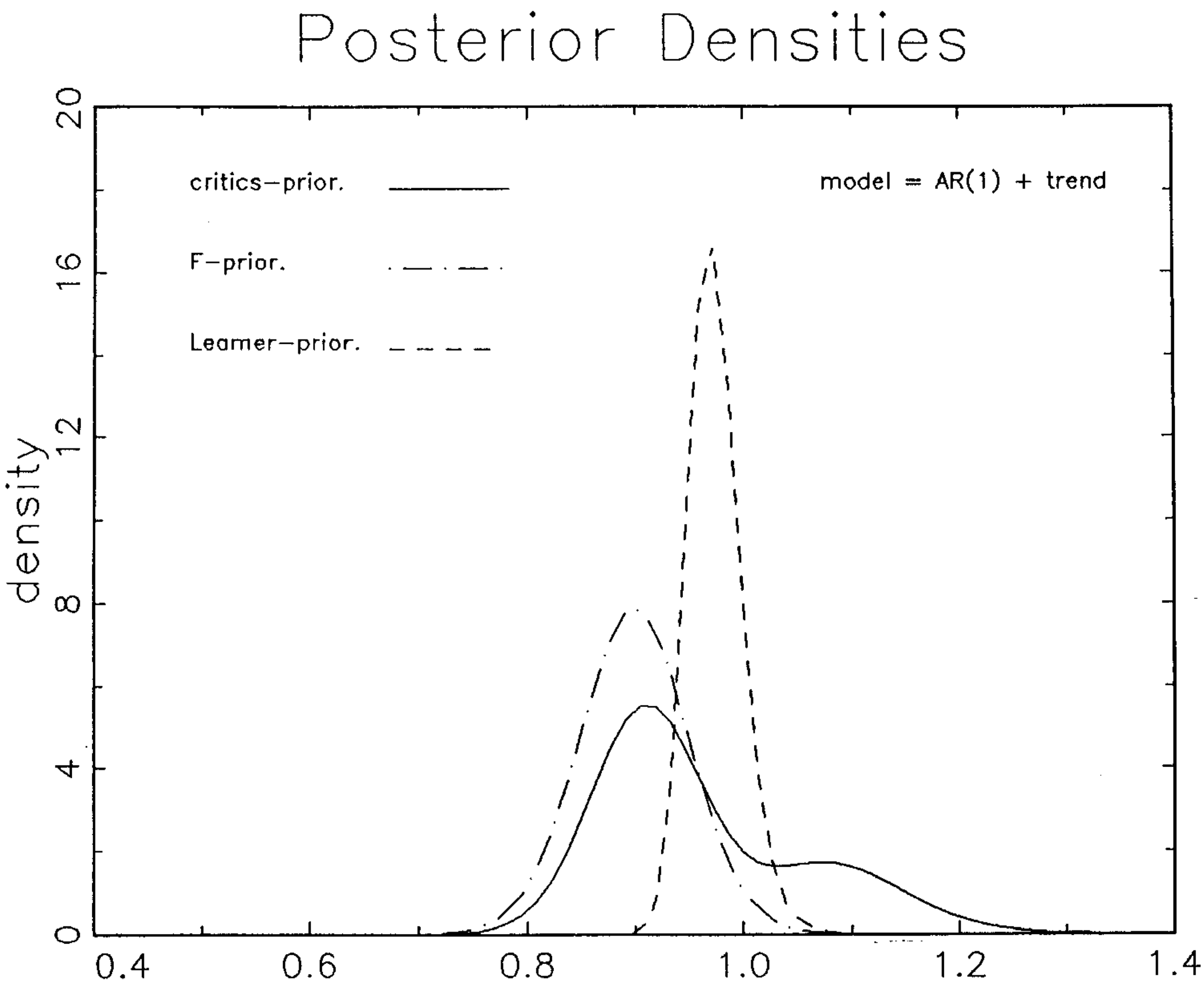

Figure 2: Real Stock Prices: 1871–1970