Discussion Paper 43

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

June 1991

ON BAYESIAN ROUTES TO UNIT ROOTS

Peter C. Schotman*
Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis

and University of Limburg

Herman K. van Dijk*
Erasmus University

ABSTRACT

This paper is a comment on P.C.B. Phillips, "To criticise the critics: an objective Bayesian analysis of stochastic trends" [Phillips (1990)]. Departing from the likelihood of an univariate autoregressive model different routes that lead to a posterior odds analysis of the unit root hypothesis are explored, where the differences in routes are due to the different choices of the prior. Improper priors like the uniform and the Jeffreys prior are less suited for Bayesian inference on a sharp null hypothesis as the unit root. A proper normal prior on the mean of the process is analyzed and empirical results using extended Nelson/Plosser data are presented.

*The research for this paper was started while the first author was visiting the Institute for Empirical Macroeconomics (IEM) in Minneapolis (MN) during the summer of 1990. We thank George Hall of the IEM for able research assistance.

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

Any opinions, findings, conclusions, or recommendations expressed herein are those of the author(s) and not necessarily those of the National Science Foundation, the University of Minnesota, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1. INTRODUCTION.

Peter Phillips' paper [Phillips (1990)] is a provocative criticism on some Bayesian contributions to the discussion of testing for unit roots. The work that Phillips reacts upon is itself a criticism on classical testing for the unit root hypothesis.[1] We are, therefore, reacting on a criticism of a criticism. Although our comments are directed at Phillips' paper, in some cases they should not be regarded as objections to his arguments, but rather as additional remarks on the work he criticizes.

Phillips' main point concerns the mechanical use of flat priors in a Bayesian analysis of time series models. He aptly demonstrates that flat priors are not uninformative but unwittingly introduce a tendency toward stationary models, and he argues that this might explain the different conclusions reached in Bayesian and classical analysis of the unit root hypothesis: *"when Bayesian and classical procedures lead to divergent conclusions we should seek first to find the answer in the prior rather than rush out to announce the failure of classical methods"* (Phillips p.2). In his paper he refrains from taking sides in the debate about the superiority of Bayesian or classical methods in general, seeing value in both approaches. The larger part of the paper (sections 3 and 4) offers an explicit and interesting Bayesian analysis of stationarity versus nonstationarity. This part of the paper is also the most provocative, since it contains detailed prescriptions how Bayesian methods should be properly applied. Our comments focus on this second part of his paper.

Phillips stresses *"the fragility of Bayesian inferences to the specification of the prior"* (Phillips, p.9). We investigate several aspects of this sensitivity within an autoregressive model with linear trend.

---

[1] Sims (1988), Sims and Uhlig (1988), DeJong and Whiteman (1989).

Departing from the likelihood function of this model we follow different routes, that diverge not only in destination but also in the paths through the jungle of priors and model representations.

As a preliminary we need some technical results on what seem minor changes in the specification of the deterministic components of the model (section 2). Paradoxically, the "ignorance" prior proposed by Phillips downweights the unit root hypothesis relative to a flat prior in a model with trend and intercept. Further, and contrary to results of DeJong and Whiteman (1989), a flat prior gives probability one to the unit root in a different parameterization of this model (section 3). The core of our comments concerns the hypothesis of interest (section 4) and its consequences for Bayesian inference. Is the relevant null an exact unit root or nonstationarity in general? In our view it is the sharp unit root null that matters. Posterior odds is the principal Bayesian tool for this problem (section 5). Their application requires an informative prior (section 6). For the empirical parts we make use of the Nelson/ Plosser data, see Nelson and Plosser (1982), extended to include the additional 18 years 1971-1988; see the data appendix. Finally, section 7 summarizes.

We concentrate on the Bayesian aspects of testing for a unit root within a univariate autoregressive model. It is beyond the scope of these comments to explore other models (*e.g.* ARIMA, fractional integration), and the sensitivity with respect to the presence of heteroskedasticity and fat-tailed error distributions. Several authors even contend that the unit root hypothesis is untestable and maybe not important at all (see Christiano and Eichenbaum (1990), Cochrane (1991)). Pragmatically, univariate autoregressive models remain suited as a preliminary step in constructing multivariate models relating variables with approximately the same type of trend, deterministic or stochastic.

2. NUISANCE PARAMETERS.

Phillips criticizes the flat prior used by the Bayesian unit rooters because of the apparent bias these priors introduce toward stationarity. He performs a simulation exercise where repeatly samples of 50 observations are generated from a pure random walk process. Then, in a simple AR(1), $y_t=\rho y_{t-1}+\varepsilon_t$, with a flat prior on ρ, the rejection frequency $E\{P_F(\rho\geq 1)\}$ is estimated as equal to 0.389. In a model with fitted intercept and trend the bias is even worse, $E\{P_F(\rho\geq 1)\} = 0.0456$ (see Phillips (eq. (12) and eq. (24))). Instead of the flat prior Phillips proposes a Jeffreys prior which has some desirable invariance properties and is "objective" or "ignorant" in a certain sense (see Phillips (p. 11-13)). The same simulation exercise establishes that the Jeffreys prior attaches appreciably larger probability to nonstationarity. The rejection frequencies increase to $E\{P_J(\rho\geq 1)\} = 0.625$ for the simple AR(1) and $E\{P_J(\rho\geq 1)\} = 0.2975$ for the AR(1) with trend and intercept. These outcomes show that the Jeffreys prior is also subject to bias, and moreover that the direction and amount of bias depend on nuisance parameters in the model. The downward bias in the model with intercept and trend occurs even though the Jeffreys prior in this model is steeply upward sloping for $\rho\geq 1$ (see Phillips (figures 1(i) to 1(iv))).

These differences warrant a closer investigation of the effect of the intercept and trend. To begin with, the prior densities plotted by Phillips are conditional priors of ρ given the nuisance parameters μ and β in

$$y_t = \mu + \beta t + \rho y_{t-1} + \varepsilon_t, \qquad (1)$$

Different values of μ and β produce different curves in the figure. The only marginal prior density of ρ is the solid line in figure 1(i), which pertains to the simple AR(1) where there are no trend and intercept.

In order to derive the marginal prior density of ρ it is convenient to write the AR(1) as

$$y_t = \gamma + \delta t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t \qquad (2)$$

which has the format of the linear regression model with AR(1) errors. The parameters in eq. (2) have a well defined meaning: δ is the mean growth rate of $\{y_t\}$; γ denotes the intercept of the deterministic linear trend $D_t = \gamma + \delta t$. If there is no trend growth (δ=0) the parameter γ is the unconditional mean of $\{y_t\}$. One of the attractive properties of the Jeffreys prior is its invariance with respect to the parameterization of the model. For the Bayesian posterior inference on ρ we can start either from the reduced form representation (1) or the "structural" representation (2). It will become apparent shortly why eq. (2) is more convenient.

To derive the Jeffreys prior we compute the expected value of the Hessian of the log-likelihood function: $I_{\vartheta\vartheta} = -E\left(\frac{\partial^2 \ell nL(\vartheta)}{\partial\vartheta\partial\vartheta'}\right)$. For model (2) we have, using similar computations as Phillips (p.18-19), and letting $\vartheta = (\gamma \quad \delta \quad \rho \quad \sigma)'$:

$$I_{\vartheta\vartheta} = \begin{pmatrix} \frac{1}{\sigma^2}(\tilde{X}'\tilde{X}) & 0 & 0 \\ 0 & I_{\rho\rho} & 0 \\ 0 & 0 & \frac{2T}{\sigma^2} \end{pmatrix}, \qquad (3)$$

where:

$$\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_T)', \text{ with } \tilde{x}_t = x_t - \rho x_{t-1} \text{ and } x_t = (1 \mid t)',$$

$$I_{\rho\rho} = \alpha_0(\rho) + \frac{(y_0 - \gamma)^2}{\sigma^2}\left(\frac{1 - \rho^{2T}}{1 - \rho^2}\right),$$

$$\alpha_0(\rho) = \frac{1}{1 - \rho^2}\left(T - \frac{1 - \rho^{2T}}{1 - \rho^2}\right), \text{ (compare Phillips (eq. (14), p. 19)).}$$

For the moment we ignore the term due to the initial condition y_0 by assuming $y_0 = \gamma$. Using (3) the Jeffreys prior then becomes

$$\pi_J(\rho,\gamma,\delta,\sigma) \propto \left(\det(I_{\vartheta\vartheta})\right)^{1/2} \propto \sigma^{-3}|\tilde{X}'\tilde{X}|^{1/2}\alpha_0(\rho)^{1/2}$$

$$\propto \sigma^{-3}(1-\rho)^2\alpha_0(\rho)^{1/2} \qquad (4)$$

The last proportionality sign can be verified by direct calculation of the determinant of the (2×2) matrix $\tilde{X}'\tilde{X}$. Eq. (4) has the convenient property that the priors on all elements of ϑ are independent. Note that $\pi_J(\rho,\gamma,\delta,\sigma)$ does not depend on γ and δ. The implied priors and γ and δ are flat, while the prior of ρ is similar but not identical to the J-prior of ρ in the simple AR(1) without trend and intercept. It differs by the additional factor $(1-\rho)^2/\sigma^2$ in eq. (4).

The nuisance parameters γ and δ affect the shape of the marginal prior of ρ. Though the interpretation of ρ does not change with the variables that we include in x_t, our prior beliefs about ρ change according to the particular deterministic components that we add to the model. The differences are most pronounced close to the unit root. From eq. (4) it follows that the Jeffreys prior for ρ drops to zero when $\rho\to1$. This follows from representation (2) of the AR model. In eq. (2) the parameter γ vanishes from the model (γ is not identified under the unit root); hence the likelihood function is flat and the information matrix $I_{\vartheta\vartheta}$ has a singularity at $\rho=1$. Figure 1 shows the Jeffreys prior for the simple AR(1), the AR(1) with a constant term, and with constant and trend. For the latter two models the Jeffreys prior gives very low probability to values of ρ close to unity. This partly explains the observed bias towards stationarity of $P_J(\rho\geq1)$ in models with a fitted intercept and trend. While the Jeffreys prior is invariant to the parameterization of the model, it is sensitive to the exact specification of the model.

FIGURE 1A: J–Priors of ρ in AR(1)

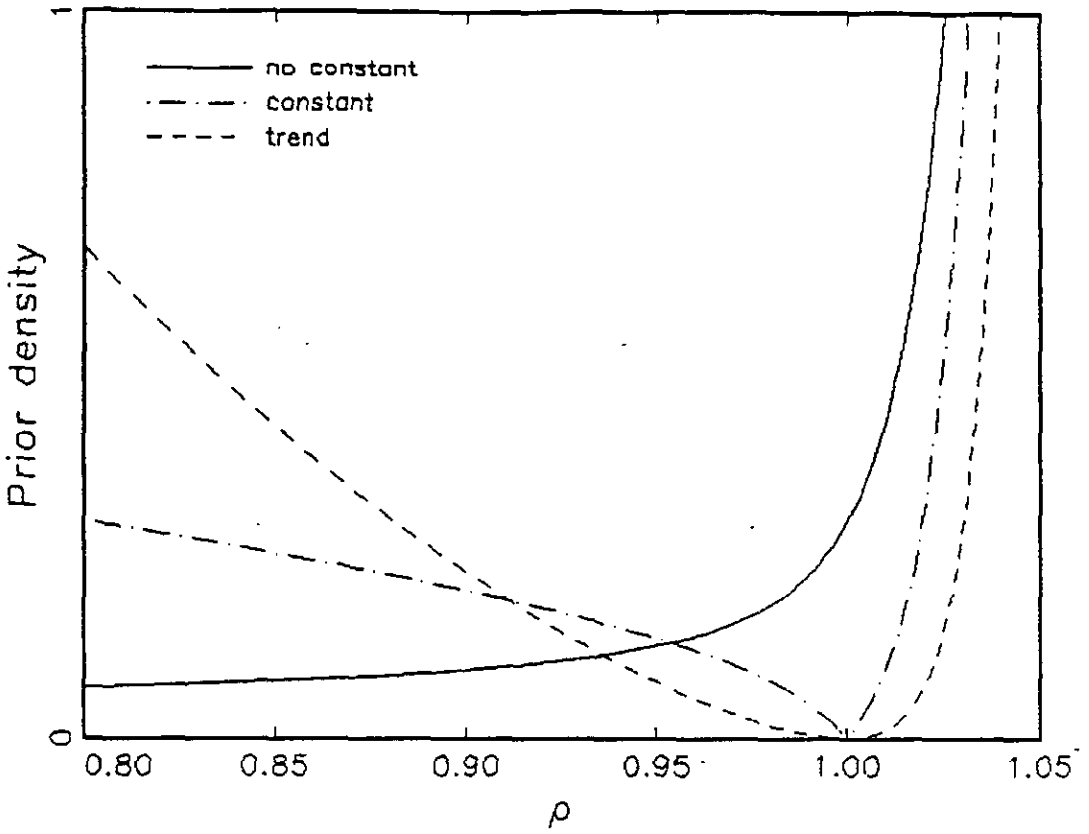
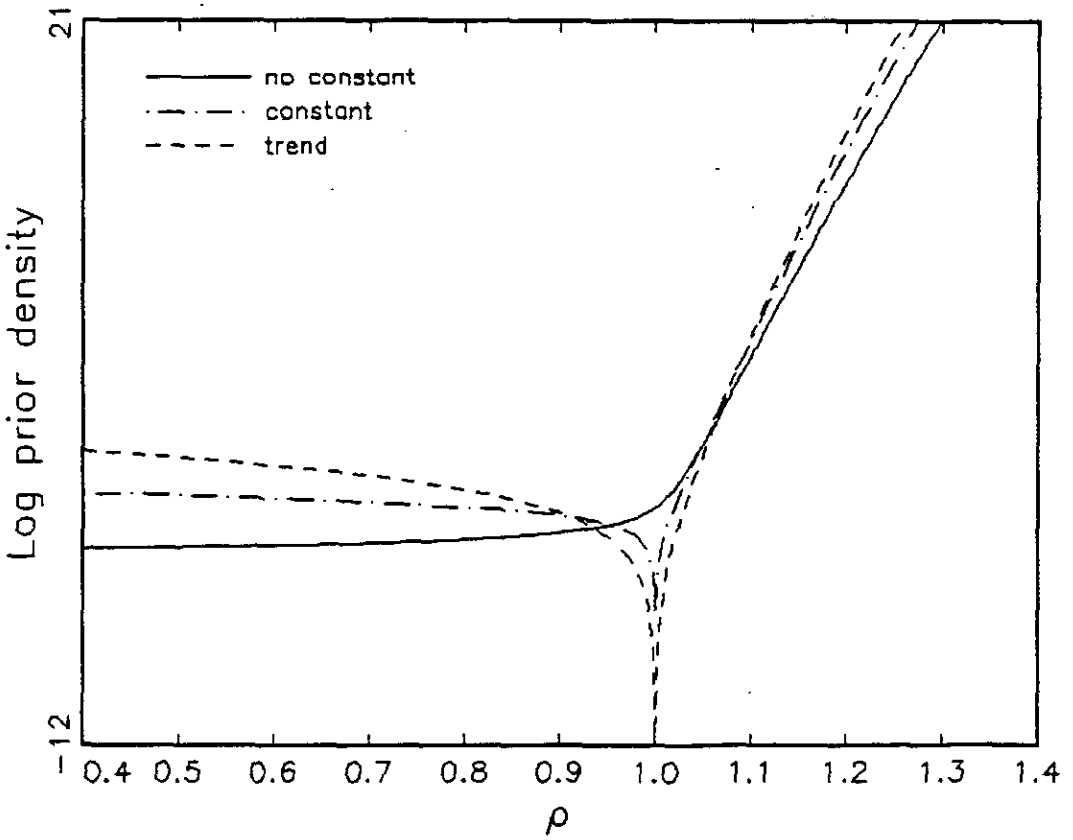


FIGURE 1B: J–Priors of ρ in AR(1)



3. THE PATHOLOGY OF FLAT PRIORS.

Phillips rightly warns against the mechanical use of a flat prior once we are out of the textbook linear regression model. His paper leaves the impression, however, that a flat prior on ρ strongly favours stationarity. But depending on the parameterization a flat prior can also have the opposite effect. Continuing the example of the AR(1) with trend parameterized as in eq. (2), a flat prior takes the form $\pi_F(\rho,\sigma,\gamma,\delta) \propto \sigma^{-1}$. Since conditional on ρ the model is linear in γ and δ, the standard integration steps for the linear model can be used to obtain the marginal posterior of $\rho \neq 1$ (see also Zellner and Tiao (1964) and Zellner (1971)),

$$p_F(\rho|Y) \propto |\tilde{X}'\tilde{X}|^{-1/2}(\hat{u}'\hat{u})^{-(T-2)/2}, \tag{5}$$

where: $\hat{u} = M_{\tilde{X}}\tilde{y}$,

$M_{\tilde{X}} = I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$,

$\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_T)'$, and $\tilde{y}_t = y_t - \rho y_{t-1}$.

From eq. (4) we know that $|\tilde{X}'\tilde{X}| \propto (1-\rho)^4$. The data matrices $\tilde{y}$ and $\tilde{X}$ depend on ρ, but the projection matrix $M_{\tilde{X}}$ does not. Due to the specific nature of the deterministic regressors, the space spanned by $\tilde{X}$ is the same space as spanned by the columns of X, and hence $M_{\tilde{X}} = M_X$, which does not depend on ρ. Since $\tilde{y}$ is linear in ρ, the "residual sum of squares" $\hat{u}'\hat{u}$ is a quadratic function of ρ. Despite the singularity of $(\tilde{X}'\tilde{X})$ at $\rho=1$ the "residual sum of squares" is well behaved as $\rho \to 1$. Close to $\rho=1$ the posterior $p_F(\rho|Y)$ is dominated by the factor $(1-\rho)^{-2}$, and hence is not integrable. A flat prior with the parameterization (2) assigns a posterior probability of one to the unit root.

The Jeffreys prior safeguards against pathological behavior of the posterior. Because it is proportional to $|\tilde{X}'\tilde{X}|^{1/2}$, it cancels the non-integrable factor in $p_F(\rho|Y)$. Applying the Jeffreys prior $\pi_J(\rho,\gamma,\delta,\sigma)$ in eq.

(4), the marginal posterior of ρ becomes

$$p_J(\rho|Y) \propto \alpha_0(\rho)^{1/2}(\hat{u}'\hat{u})^{-T/2}, \tag{6}$$

which is a proper density. It is equal to Phillips' equation (22) that he obtained using the Laplace approximation. We could derive the posterior exactly because we treated the initial observation differently, $y_0=\gamma$ instead of $y_0=0$. The role of the initial conditions will come up again in section 6.

Figure 2 gives an example of the empirical implications of different priors for the Nelson / Plosser series "Stock prices" (1871-1988). The figure shows three different marginal posterior densities of ρ for an AR(1) with trend and intercept. The flat prior used by DeJong and Whiteman (1989), labelled DJW, is most to the left and gives lowest probability to nonstationary alternatives. The Jeffreys prior attaches considerable probability mass to the nonstationary region. The flat prior of the "structural" model has an asymptote at $\rho=1$. Marginal posteriors for the other Nelson / Plosser series reveal a similar ranking among the posteriors.

4. THE HYPOTHESIS OF INTEREST.

In the empirical examples it appears that the evidence against stationarity stems largely from the fact that the "objective" prior assigns enormous probability mass to highly explosive models which often results in a bimodal posterior where one of the modes is in the explosive region of the parameter space. See figures 4(i) to 4(xiv) of Phillips (1990). In his table 4 Phillips presents the posterior probability of $\rho>1$, and interprets these probabilities as evidence of stochastic nonstationarity.

Stochastic nonstationarity is, however, not the same thing as a unit

FIGURE 2: Stock prices

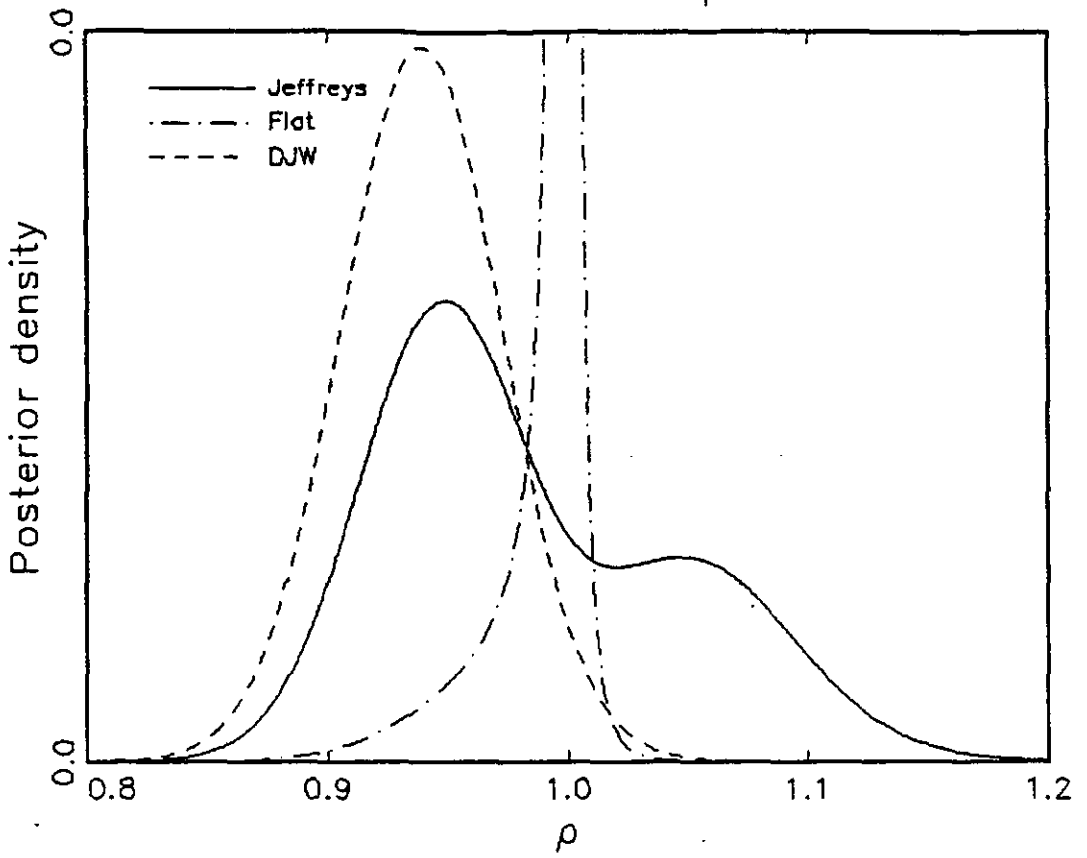
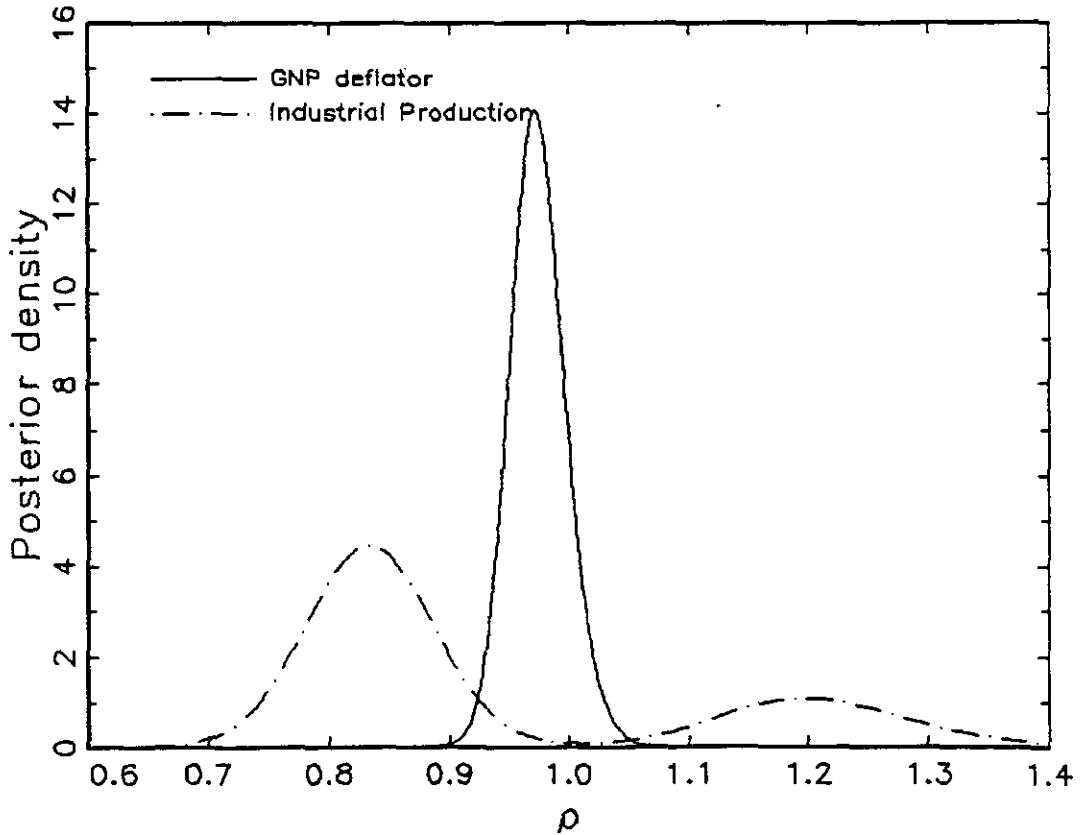


FIGURE 3: Unit root or nonstationarity?



root. In many applications the economic theory leads to models with a unit root null hypothesis. Phillips (1990, p.4-5) provides some of the leading examples: the efficient market hypothesis, or the permanent income model. Confronted with a posterior mode far above one, why should one interpret this as evidence in favour of a unit root? Why is a sharp mode located around some $\rho>1$ more evidence of unit root behaviour than a mode located slightly below $\rho=1$? To illustrate the point, consider the two posteriors in figure 3.[2] The first, unimodal, posterior *U* has its mode slightly below one while the second posterior *B* is bimodal with the modes rather different from unity. The sharply peaked posterior has almost all its probability mass below $\rho=1$, and leads to a high probability $P(\rho<1)$. The bimodal posterior assigns a high probability to the nonstationary region. Yet an exact unit root is more probable with posterior *U* than with posterior *B*.

Before going into the details of appropriate priors one should first state with what purpose the analysis is conducted, and state the hypothesis of interest. It seems that in the literature two cases are not always clearly distinguished. First there is the unit root hypothesis versus the alternative of stationarity:

$$H_0: \rho = 1 \qquad \leftrightarrow \qquad H_1: |\rho| < 1.$$

But Phillips (1990, p. 27, table 4) appears to be testing

$$H_0: \rho \geq 1 \qquad \leftrightarrow \qquad H_1: \rho < 1.$$

The relevant pair of hypotheses depends on the question being asked. The economic examples referred to above deal with the exact unit root hypothesis, not with the question whether a root is larger or smaller than unity. In time series modelling one of the first steps consists of transforming the data to

---

[2] The two posteriors in the figure have been obtained from an AR(3) model for the extended Nelson / Plosser series "GNP deflator" and "Industrial Production".

a series that is stationary, usually by taking first differences. To find out whether this transformation is required was the primary motivation of Nelson and Plosser (1982), who attempted to determine whether a series was trend stationary (TS) or difference stationary (DS). Within an autoregressive representation the trend and cyclical components of TS and DS look completely different. We emphasize these stylized facts in figure 4 for real GNP per capita in the U.S. (1909-1988). Figure 4a shows the actual data and the deterministic linear trend, estimated with an AR(3) model in levels; figure 4b shows a stochastic trend implied by an AR(2) in first differences. Empirically a stochastic trend, which is a random walk with drift, can account for almost all fluctuations in U.S. real GNP. As a result figure 2c suggests that there was no major business cycle in the thirties.[3] Although the current decomposition can be sensitive to adding MA components and although AR models have some limitations, the figure dramatizes the effect of a unit root, when all its implications are taken seriously.

The relevant pair of hypotheses for this problem is the first one, involving the exact unit root. The definition of a stochastic trend is confined to DS series, and does not extend to an explosive model.[4] First differencing a time series with a root of 1.5 is not an effective transformation to induce stationarity. The unit root hypothesis is a prime example of a sharp null hypothesis. Phillips (1988, p.348) offers additional forceful

---

[3] The estimated deterministic trendline is (standard errors in parenthesis):

$$D_t = \underset{(0.07)}{6.98} + \underset{(0.001)}{0.020}\ t$$

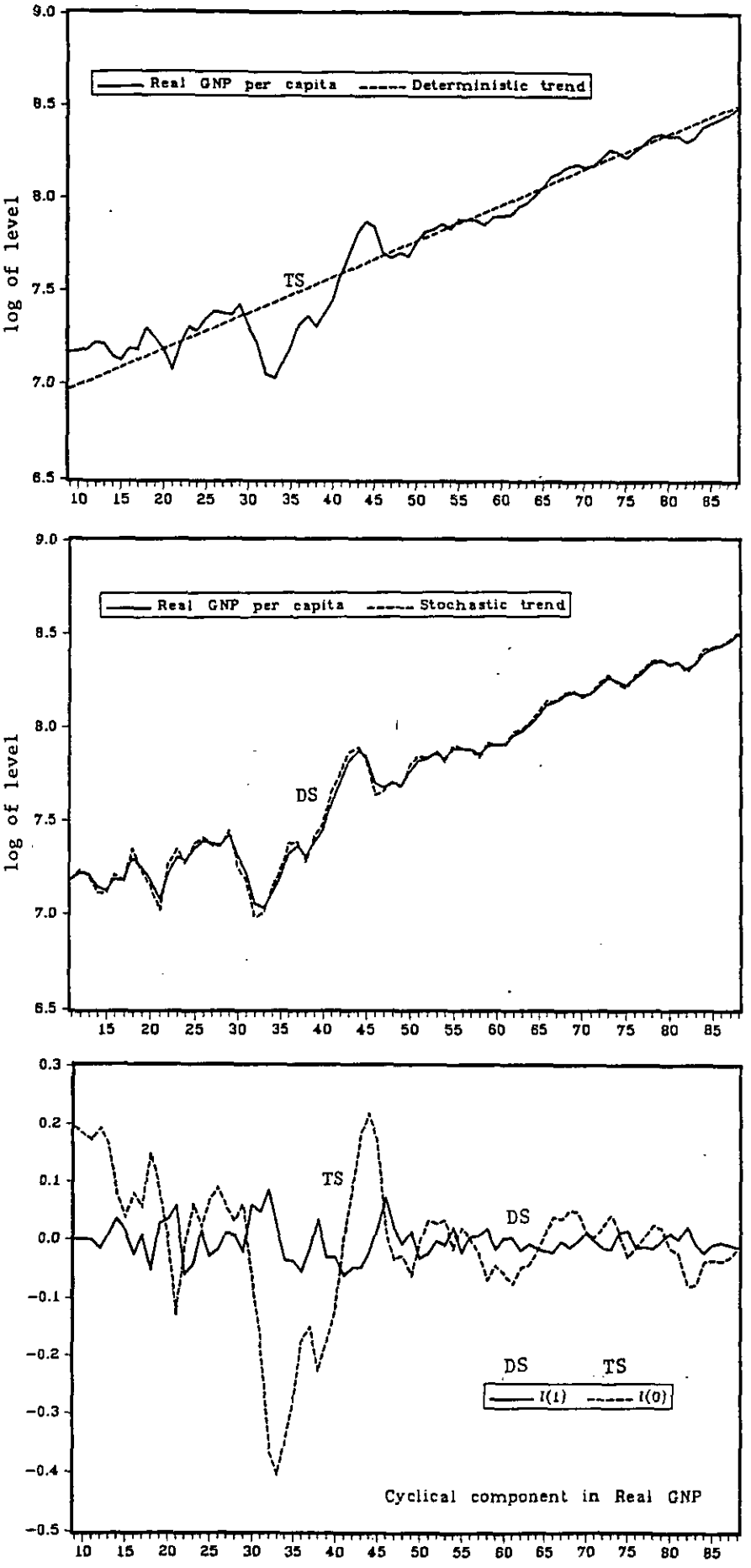
For the stochastic trend the Beveridge and Nelson (1981) decomposition of the ARI(2,1) gives

$$D_t = D_{t-1} + \underset{(0.009)}{0.017}\ t + \underset{(0.25)}{1.35}\ \varepsilon_t$$

Alternative trend/cycle decompositions of course exist (see Harvey (1990)).

[3] See, however, Sims (1989) for a operational definition of a trend in models with an explosive AR root.

Figure 4. Trend-cycle decompositions of US real GNP, 1909-1988



arguments of this statement. The second pair of hypotheses seems more appropriate in general misspecification tests. Stationarity is doubted without a clear alternative for some data series. Unlike the exact unit roots, explosive roots do not have the profound implications that have led to the development of cointegration and stochastic trends, or the concern with persistent versus transitory shocks in macroeconomics. The econometrician's response to explosive roots is respecification of the model: possibly nonlinear trends, maybe a log transformation is required.

5. TESTING A SHARP NULL WITH AN IMPROPER PRIOR.

Bayesian methods are particularly suited for the decision problem encountered in unit root econometrics. The unit root hypothesis is often an auxiliary hypothesis about the formulation of a dynamic model: TS or DS. There is not always an economic reason to prefer one over the other. The unit root issue is important for correct statistical inference (see, *e.g.* Phillips (1987), Durlauf and Phillips (1988)), and affects the economic interpretation of shocks (see Campbell and Mankiw (1987) or Blanchard and Quah (1990)), but it is a choice problem rather than a testing problem. The loss involved with type I and type II errors might be equal, unlike the preoccupation with the type I error in classical hypothesis testing. Second, the unit root test in autoregressive models compares a point hypothesis (ρ=1) with a composite alternative (ρ<1). Such a test can only be formulated in a classical framework if the point hypothesis is taken as the null. A Bayesian analysis allows a symmetric treatment of the null and the alternative.

The principal Bayesian tool for comparing hypotheses is the posterior odds ratio $K_1 = P(H_0|\text{data})/P(H_1|\text{data})$. For the second pair of hypotheses the posterior odds ratio is equivalent to computing $P(\rho \geq 1)$, since $P(\rho<1) = 1 - P(\rho \geq 1)$. In order to apply the posterior odds to a test of the sharp null

of an exact unit root against the alternative of stationarity, $|\rho| < 1$, we must assign a discrete probability to the event $\rho = 1$. Leamer (1978, sec. 4.3) and Zellner (1971, p. 297-299) provide the details of the computations. Treating H_0 and H_1 symmetrically, *i.e.* putting $P(H_0) = P(H_1)$, the posterior odds are equal to

$$K_1 = \frac{\int_{\Omega} \pi_2(\vartheta) \cdot L(Y|\vartheta,\rho=1)\, d\vartheta}{\int_{-1}^{1}\int_{\Omega} \pi_1(\rho|\vartheta) \cdot \pi_2(\vartheta) \cdot L(Y|\vartheta,\rho)\, d\vartheta\, d\rho}, \tag{7}$$

where $\pi_2(\vartheta)$ is the marginal prior of the auxiliary parameters $\vartheta \in \Omega$ in the model; $\pi_1(\rho|\vartheta)$ is the conditional prior of ρ; and $L(Y|\vartheta,\rho)$ is the likelihood function.

We now apply the posterior odds formula to the AR(1) with trend and intercept. The marginal Jeffreys prior of the auxiliary parameters $\pi_2(\sigma,\gamma,\delta) \propto \sigma^{-3}$ is improper (see eq. (4)). Since the conditional likelihood function for $\rho=1$ does not depend on γ (because γ is not identified at $\rho=1$), the integral in the numerator of (7) diverges. The integral in the denominator of the posterior odds is well defined, since

(i) the prior $\pi_1(\rho|\vartheta) = (1-\rho)^2\alpha_0(\rho) / \int_{-1}^{1}(1-\rho)^2\alpha_0(\rho)d\rho$ is proper and independent of ϑ,

(ii) $\pi_1(\rho)\int_{\Omega}\pi_2(\vartheta)\cdot L(Y|\vartheta,\rho)d\vartheta$ is proportional to the marginal posterior $p_J(\rho|Y)$ in eq. (6).

The fact that γ is locally unidentified at $\rho=1$ leads to the anomalous result that the posterior odds diverge to infinity and thus imply probability one for the unit root. Schotman and Van Dijk (1991[a]) encountered the same pathology for the flat prior on ρ and γ. Any proper prior on ρ defined on the stationary interval $\rho\in(-1,1)$ combined with an improper prior on γ will yield this result. Thus the Jeffreys prior is ill suited for a Bayesian test of the unit root hypothesis. The intuition is that the Jeffreys prior is improper on

nuisance parameters that enter the model in an asymmetric way: not identified under the null, but identified under the alternative. For more discussion on the use of improper priors we refer to Zellner and Siow (1980) and the discussion of that paper by Jaynes (1980).

6. CHOICE OF PRIOR.

The need for an informative prior greatly complicates a Bayesian procedure, if no natural reasonable informative prior is available. The analysis in the previous section suggests that for a proper posterior odds test we need to break the prior independence between γ and ρ inherent in the Jeffreys and in the flat prior. If a time series is stationary, there exists a linear deterministic trend with intercept γ. The more precise we can locate γ, the more convinced we can be that the trendline is stable and does not wander stochastically, as it does in the presence of a stochastic trend (ρ=1). On the other hand, if the position of the trendline (γ) is not well determined, this is an indication it may shift over time, which suggests unit root behaviour. An alternative formulation of the unit root hypothesis is the question: "How much can we learn about the level of a series?". The only parameter that conveys information about the level of the series is γ. The parameters γ and ρ are tightly connected, γ not really being a nuisance parameter.

A technical solution of the problem of specifying a prior on γ that leads to properly behaved posterior odds is to exploit the information in the initial condition of the time series. So far all the results used the conditional likelihood function, given the initial observation y_0. Schotman and Van Dijk (1991[a,b]) suggest a normal prior on γ, centered around y_0, and with a variance conditional on ρ and σ.

$$p(\gamma|\rho,\sigma,y_0) \sim \text{Normal}\left(y_0 \, , \, \frac{\sigma^2}{1-\rho^2}\right) \tag{8}$$

The variance equals the unconditional variance of an AR(1) process. This prior is proper and weakly informative. It is only defined for $|\rho|<1$. It tends to a flat prior if $\rho \to 1$. Combined with a proper prior on ρ this leads to well defined posterior odds.[5] In Schotman and Van Dijk (1991[a,b]) we used a flat prior on ρ over a part of the stationary region: $\rho \in [\alpha,1)$. The formula for posterior odds (eq. (7)) becomes

$$K_1 = \frac{p(\rho=1|Y)}{\frac{1}{1-\alpha}\int_{\alpha}^{1} p(\rho|Y)d\rho}, \tag{9}$$

where $p(\rho|Y)$ is the marginal posterior of ρ for $\rho \in [\alpha,1)$, and $p(\rho=1|Y) = \lim_{\rho \to 1} p(\rho|Y)$. The procedure amounts to comparing the value of the posterior in $\rho=1$ to the average value of the posterior under the alternative $\rho \in [\alpha,1)$.

Application of the posterior odds ratio has serious effects for the empirical results. In table 1 we compare the probability of nonstationaruty $P_J(\rho>1)$, obtained with the Jeffreys prior, with the probability of an exact unit root $Pr(\rho=1)$, using the normal prior on γ and setting α equal to 0.8. For further reference we also report the classical Augmented Dickey Fuller test. These three different quantities are computed with two data sets: the original Nelson / Plosser data, and the extended data set.

The posterior odds ratios reveal that the data for most series are not very informative about the unit root hypothesis. The posterior probability of a unit root is between .25 and .75 for 9 of the 14 series in the "1970"

---

[5] Technically, we use the so-called exact likelihood instead of the conditional likelihood.

sample. "Unemployment" is the only clear stationary series, in close agreement with the classical test. The integrated series are "Consumer prices", "Velocity", "Interest rate" and "Stock prices". The latter results are qualitatively the same as those of Phillips (1990). These four series have a high probability of nonstationarity, relative to the general tendency

TABLE 1: *Posterior probabilities of stochastic nonstationarity*

	1970 sample			1988 sample		
Series	$P_J(\rho \geq 1)$	ADF	$Pr(\rho=1)$	$P_J(\rho \geq 1)$	ADF	$Pr(\rho=1)$
Real GNP	0.012	-2.99	0.365	0.005	-3.46	0.309
Nominal GNP	0.074	-2.32	0.567	0.131	-2.02	0.752
Real GNP per cap.	0.010	-3.05	0.347	0.004	-3.52	0.223
Industrial Prod.	0.188	-2.53	0.283	0.263	-2.66	0.286
Employment	0.040	-2.66	0.510	0.034	-2.87	0.614
Unemployment	0.086	-3.55	0.169	0.063	-3.92	0.110
GNP deflator	0.020	-2.52	0.613	0.134	-1.59	0.884
Consumer prices	0.176	-1.97	0.883	0.449	-1.20	0.952
Wages	0.045	-2.23	0.588	0.060	-2.12	0.751
Real wages	0.014	-3.05	0.517	0.555	-1.68	0.756
Money stock	0.008	-3.08	0.407	0.021	-2.86	0.567
Velocity	0.537	-1.66	0.754	0.364	-1.60	0.861
Interest rate	0.996	0.39	0.854	0.650	-0.54	0.849
Stock prices	0.215	-2.12	0.935	0.313	-1.92	0.968

The first column is adapted from Phillips (1990). ADF is the Augmented Dickey Fuller test $\hat{\tau}_\tau$, see Dickey and Fuller (1979, 1981) and Fuller (1976); the second column is from Nelson and Plosser (1982). Alternative region in exact unit root test is [0.8, 1). All results are based on AR(3) model, except for "unemployment" which is AR(4).

toward stationarity in the first column of table 1. The additional data for the seventies and eighties provide a surprisingly amount of extra information on the unit root hypothesis. For the "1988" sample we have 8 series that appear I(1), and 2 series I(0), the second being the much debated variable "Real GNP per capita". For the two stationary series we find again a close correspondence with the classical test: the ADF rejects a unit root in "Real GNP per capita" at the 5% level. As a general pattern the posterior probability $Pr(\rho=1)$ indicates that the real variables are I(0) while the nominal and price series are in most cases I(0). This is in contrast to the results of the Dickey Fuller for the "1970" sample where I(1) can not be rejected except for Unemployment. The extra data in the "1988" sample are informative in the sense that now I(1) is rejected for three series.

In many cases the differences between the results of Phillips (1990) and our results are not due to the different priors that are being used, but more to the hypothesis that is tested. Figures 5 and 6 illustrate this point. The figures compare the posteriors implied by the Jeffreys prior ($p_J(\rho|Y)$) with the posterior implied by the normal prior on γ ($p_N(\rho|Y)$). Comparing figures 5A and 5B, we see that increasing the sample size for "Real GNP per capita" has the same effect on the marginal posterior of ρ for both priors: a tendency toward stationarity. Figures 6A and 6B show the analogous pattern for "Consumer prices": here a strong tendency toward $\rho=1$.

Finally, our flat prior on ρ over a small and fixed range might not be the best choice of prior for ρ. A sensitivity analysis along the lines of Schotman and Van Dijk (1991[a]) can yield insights whether this has serious consequence for the posterior odds test. Some preliminary exercises indicate that the qualitative conclusions on the relative probabilities of I(0) and I(1) are not much affected, though the actual computed probabilities are different.

FIGURE 5A: Real GNP per cap. (J−prior)

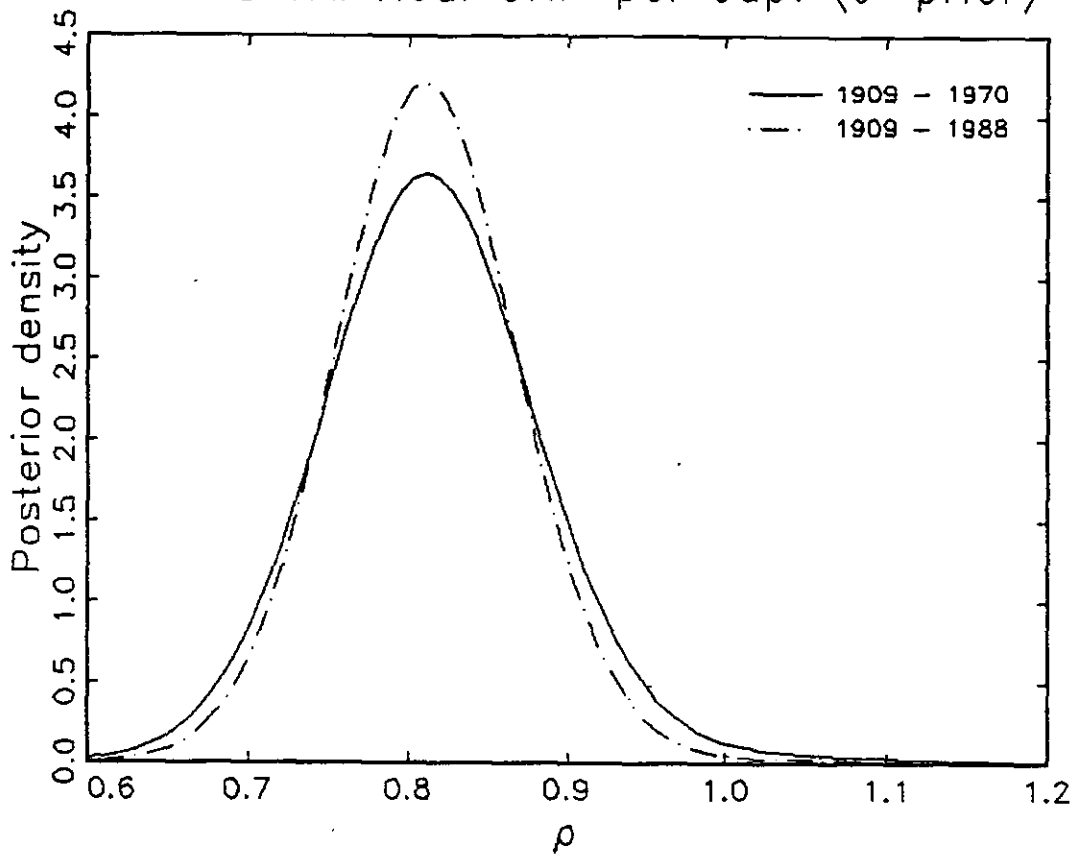


FIGURE 5B: Real GNP per cap. (N−prior)

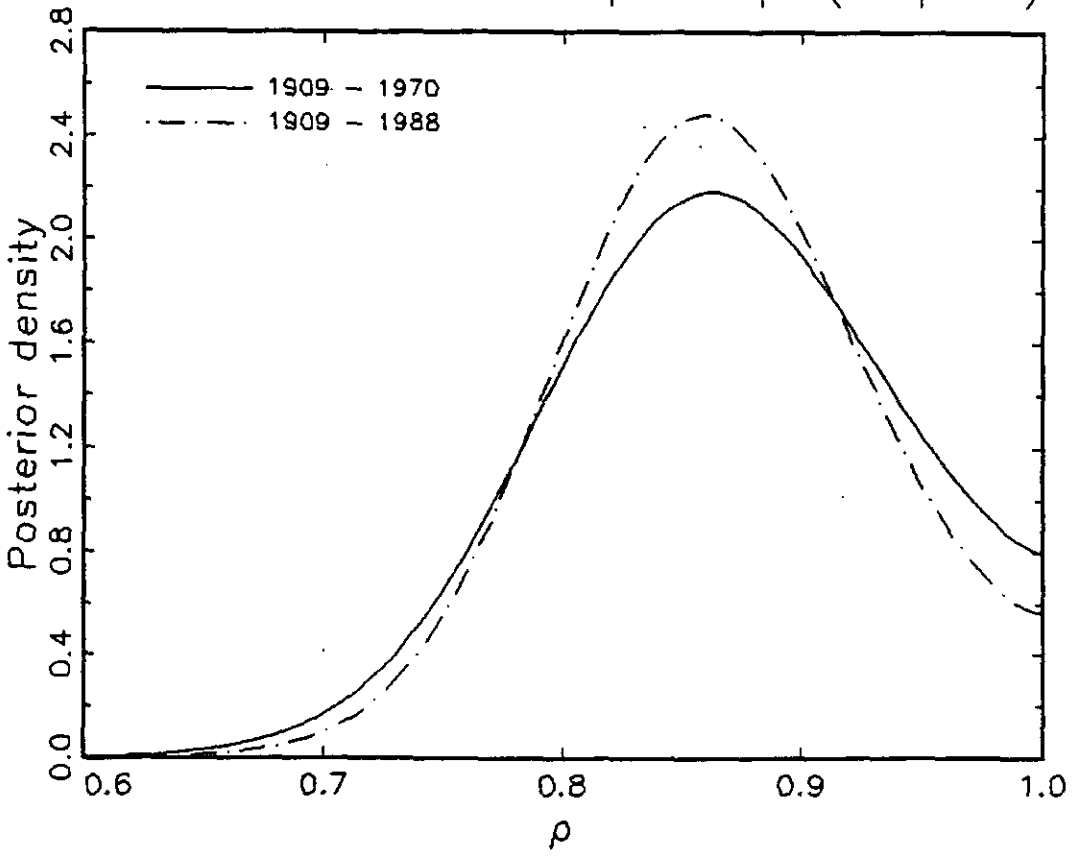


FIGURE 6A: Consumer prices (J–prior)

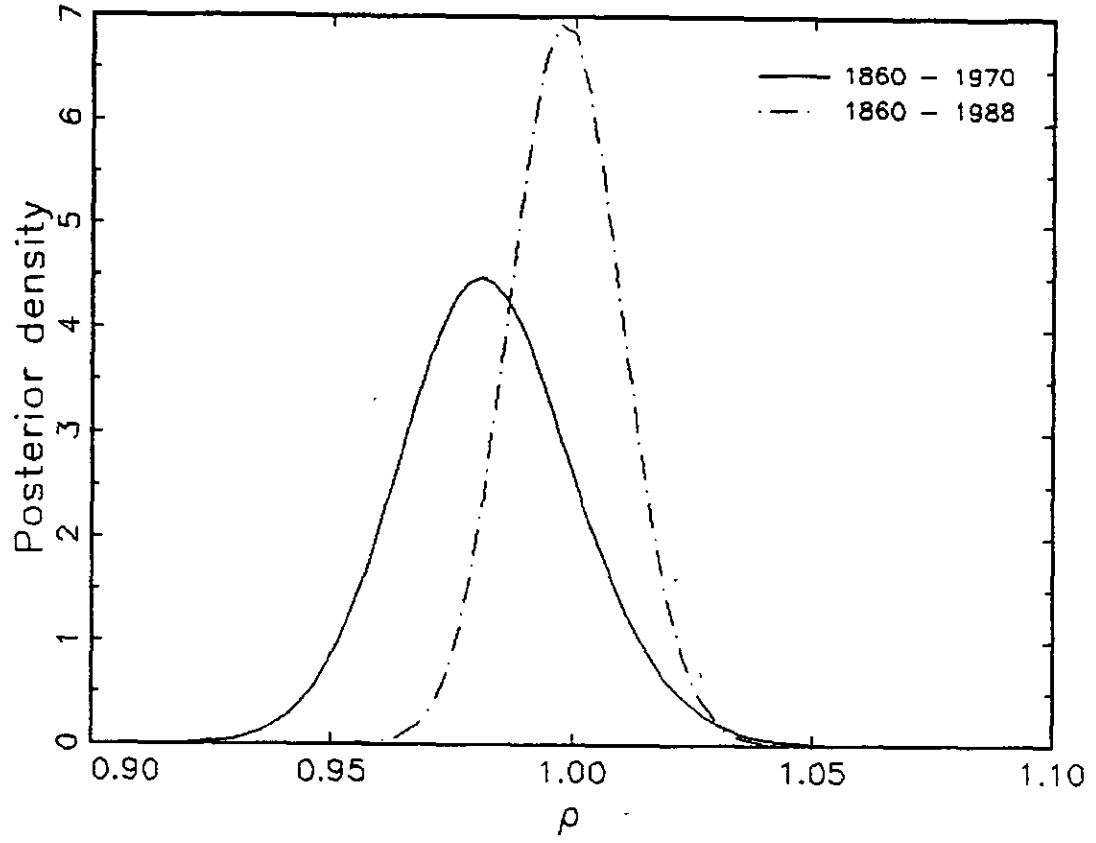
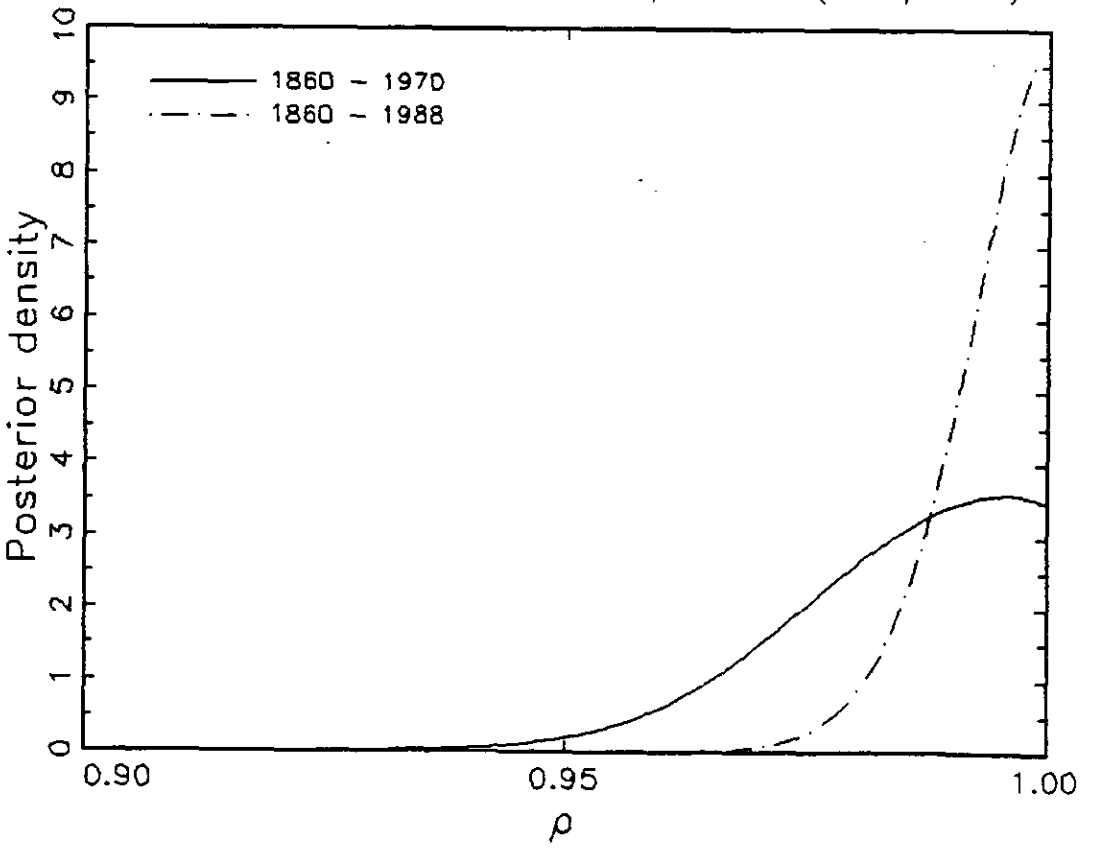


FIGURE 6B: Consumer prices (N–prior)



7. CONCLUDING REMARKS.

The Jeffreys prior has as advantage in the analysis of autoregressive time series models that it is invariant with respect to the parameterization. But we do pay a price for the invariance. The J-prior depends on the sample size, the data, and the complete structure of the model. As shown in section 2, especially our prior beliefs about ρ change according the the specification of the deterministic components in the model. Further, as argued in section 4, the whole point of the Bayesian exercise is in discriminating between two well-defined hypotheses. This problem requires an informative proper prior, if only because of the special importance of the point ρ=1. "Objective" Bayesian methods for testing a sharp null hypothesis do not exist (see Berger and Delampady (1987)).

Finally, the preposition "on" in the title has several interpretations. We have used the interpretation: "about" or "concerning" and attempted to explain the implications of the choice of different priors for inference on unit roots. We made extensive use of graphical, analytical and some numerical techniques. An other more literal interpretation of the preposition "on" is "to march on". Peter Phillips paper is also attractive since studying it may induce other researchers to explore Bayesian routes in more detail.

APPENDIX : Extension of Nelson / Plosser dataset.

The original time series of Nelson and Plosser (1982) are annual observations ending in 1970. We extended all 14 series to 1988. Series are defined as:

Real GNP	: (A), Appendix II.
Nominal GNP	: (A), Appendix II.
Real GNP per cap.	: = Real GNP / Population, (B), Population: Total.
Industrial Production	: (A), Indexes, total.
Employment	: (A), Civilian Labor Force + Resident Armed Forces.
Unemployment rate	: (A), Total unemployment rate civilian workers.
GNP Deflator	: (A), Implicit price deflator, Appendix II.
Consumer Prices	: (A), all items, urban consumers.
Wages	: (A), Average weekly earnings of production workers.
Real wages	: = Wages / Consumer Prices.
Money Stock	: (C), M2, definition of 1980.

Velocity : = Nominal GNP / (Currency plus Demand Deposits) (C).
Interest rate : (C), yield offered on a recently offered A-rated utility bond.
Stock Prices : (A), Standard and Poor 500 composite stock index.

Primary data sources are:
(A) Business Statistics 61-88, U.S. Department of Commerce, Bureau of Economic Analysis.
(B) Statistical Abstract, Bureau of the Census.
(C) Banking and Monetary Statistics, Federal Reserve Bulletin.
Data are spliced to the original series. The conversion factor is the average ratio of the old series over the new series for the period 1965 - 1970.

REFERENCES.

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

Beveridge, S., and C.R. Nelson (1981), A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', Journal of Monetary Economics, 7, 151-174.

Blanchard, O. J., and D. Quah (1990), The dynamic effects of aggregate deamnd and supply shocks, American Economic Review, 79, 655-673.

Campbell, J.Y., and N.G. Mankiw (1987), Are output fluctuations transitory, Quarterly Journal of Economics, 102, 857-880.

Christiano, L.J., and M. Eichenbaum (1989), Unit roots in real GNP: do we know and do we care?, Discussion paper 18, Institute of Empirical Macroeconomics, Minnesota.

Cochrane, J.H., (1991), 'A critique of the application of unit root tests', Journal of Economic Dynamics and Control, 15, 275-284.

DeJong, D.N., and C.H. Whiteman (1989), Trends and random walks in macroeconomic time series: a reconstruction based on the likelihood principle, University of Iowa working paper.

Dickey, D.A., and W.A. Fuller (1979), Distribution of estimators for autoregressive time series with a unit root, Journal of the American Statistical Association, 74, 427-431.

Dickey, D.A, and W.A. Fuller (1981), Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica, 49, 1057-1072.

Durlauf, S.N., and P.C.B. Phillips (1988), Trends versus random walks in time series analysis, Econometrica, 56, 1333-1354.

Fuller, W.A., (1976), Introduction to statistical time series, Wiley (New York).

Harvey, A.C., (1990), Forecasting, structural time series models and the Kalman filter, Cambridge University Press (Cambridge).

Jaynes, E.T., (1980), 'Discussion', in J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.), Bayesian Statistics, University Press (Valencia, Spain), 618-629.

Leamer, E.E., (1978), Specification Searches, Wiley (New York).

Nelson, C.R., and C.I. Plosser (1982), 'Trends and random walks in macroeconomic time series', Journal of Monetary Economics, 10, 139-162.

Phillips, P.C.B., (1987), Time series regression with a unit root, Econome-

trica, 55, 277-302.

Phillips, P.C.B., (1988), 'Reflections on econometric methodology', Economic Record, 64, 344-359.

Phillips, P.C.B., (1990), 'To criticize the critics: An objective Bayesian analysis of stochastic trends', Manuscript, Cowles Foundation for Research in Economics, forthcoming in Journal of Applied Econometrics.

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

Schotman, P.C., and H.K. van Dijk (1991[b]), 'Posterior analysis of possibly integrated time series with an application to real GNP', to appear in:E. Parzen (ed.), New Directions in time series analysis, Springer Verlag (Berlin).

Sims, C.A., (1988), Bayesian skepticism on unit root econometrics, Journal of Economic Dynamics and Control, 12, 463-474.

Sims, C.A. (1989), 'Modelling trends', Discussion Paper 22, Institute for Empirical Macroeconomics, Minneapolis (MN).

Sims, C.A., and H.A. Uhlig (1988), 'Understanding unit rooters: a helicopter tour', Discussion Paper 4, Institute for Empirical Macroeconomics, Minneapolis (MN), to appear in Econometrica.

Watson, M., (1986), Univariate detrending methods with stochastic trends, Journal of Monetary Economics, 18, 17-33.

Zellner, A., (1971), An introduction to Bayesian inference in econometrics, Wiley (New York).

Zellner, A., and A. Siow (1980), 'Posterior odds ratios for selected regression hypothesis', in J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.), Bayesian Statistics, University Press (Valencia, Spain), 583-603.

Zellner, A., and G.C. Tiao (1964), 'Bayesian analysis of the regression model with autocorrelated errors', Journal of the American Statistical Association, ??, 763-778.