Journal of Economic Dynamics and Control 12 (1988) 463–474. North-Holland

BAYESIAN SKEPTICISM ON UNIT ROOT ECONOMETRICS*

Christopher A. SIMS

University of Minnesota, Minneapolis, MN 55455, USA

Received January 1988

This paper examines several grounds for doubting the value of much of the special attention recently devoted to unit root econometrics. Unit root hypotheses are less well connected to economic theory than is often suggested or assumed; distribution theory for tests of other hypotheses in models containing unit roots are less often affected by the presence of unit roots than has been widely recognized; and the Bayesian inferential theory for dynamic models is largely unaffected by the presence of unit roots. The paper displays an example to show that when Bayesian probability statements and classical marginal significance levels diverge as they do for unit root models, the marginal significance levels are misleading. The paper shows how to carry out Bayesian inference when discrete weight is given to the unit root null hypothesis in a univariate model.

1. Introduction

There seem to be two interacting seeds generating the recent surge of work on unit root inference in econometrics. One is the idea that the question of whether or not economic time series fit models with unit roots is closely linked to important questions about economic behavior. The other is the idea that the presence of unit roots in the data greatly complicates statistical inference, making it important to work out new, correct procedures to apply to this case. Both these seeds are essentially sterile ideas.

This paper begins with brief remarks on why unit roots do not deserve to be discretely singled out on the basis of reasoning about economic behavior. Then we point out the implications of recent work on the classical asymptotic distribution theory for unit roots models, which suggest that much of the applied work using data which have been differenced and transformed to achieve stationarity is misguided. Finally, the main contribution of the paper is to point out that when unit roots are present Bayesian and classical approaches to inference diverge substantially. The Bayesian procedures are simpler and more reasonable, suggesting that inference in the presence of unit

*This research was supported by NSF Grant SES 8608078.

roots is not so sharply different from inference in stationary models as the classical theory implies.

2. Behavioral foundations for unit root hypotheses

While it is easy to write down behavioral economic models which imply that ARMA statistical models fit to the data generated by the behavioral models would have unit roots, for every such model there is another with essentially the same implications for policy and for the interpretation of behavior in which the implied ARMA models are stationary. Some recent work, for example, puts forward as 'economic theory implying unit roots' models including an arbitrary postulate that technological change, an unobservable stochastic process about which economic theory has nothing to say (at least in these models), is a process with a unit root. No argument is presented that this model is more plausible behaviorally or even substantively very different from one in which technological change is stationary but with what Granger (1969) called the 'typical spectral shape' of an economic time series – much more power at low than at high frequencies.

The best known example of a behavioral theory implying unit roots is the efficient markets hypothesis, which in its most straightforward form implies that asset prices are random walks. But it has been understood since Lucas (1978) that the competitive market mechanism does not produce random walk asset price behavior, and my own paper (1984) shows that the intuition behind the efficient markets hypothesis actually applies only as an approximation at small time intervals. The theory actually implies only models with roots close to one, and the degree of their closeness will tend to vary across applications in which data are collected at different intervals. The same sort of remarks apply to Hall's (1978) random walk hypothesis for consumption. It is easy to see that the theory implies a unit root for consumption only under special assumptions, including constant real interest rates, and that in a full general equilibrium consumption will at most be random-walk-like over short spans of data. Furthermore, even this approximate conclusion depends on arbitrary separability assumptions on dynamic utility which are justified mainly by algebraic convenience and professional inertia, not by experimental evidence or intuitive plausibility.

3. Anomalies in classical inference for unit roots

Inference in economic models should not ignore the possibility that the data are nonstationary, but the attempt to apply asymptotic distribution theory allowing for nonstationarity has been in most instances wrongheaded and unenlightening. Asymptotic theory shows [see Sims, Stock and Watson (1986)] that in linear AR models with some unit roots all linear combinations of

coefficients have the same asymptotic distribution, when normalized by $\sqrt{T}$, as if the model is estimated after transformation to stationary form with a priori knowledge of the location of unit roots. (Some linear combinations have a degenerate distribution when normalized by $\sqrt{T}$; these are the ones which would be fixed a priori in transforming the model to stationary form.) Probably the most common source of concern about unit roots on the part of applied econometricians is the notion that if they do not transform a model to stationary form, the usual distribution theory does not apply. The recent results show that this concern is completely misplaced. The *only* test statistics for which asymptotic distribution theory is simple after transformation to stationary form are those statistics which would have a simple asymptotic distribution theory even if the data were not transformed. In other words, any hypotheses which can be tested after the model is transformed, can be tested with exactly the same distribution theory using the untransformed model. There is no justification for preliminary differencing or application of cointegration transformations in the belief that these steps are necessary to allow use of the usual statistical tests.

This is not to deny that the asymptotic distribution theory for multivariate models with unit roots is hard or that it may raise unavoidable difficulties for classical inference in such models. Sims, Stock and Watson (1987), for example, show that the distribution theory of tests for block Granger casual priority in autoregressive models depends on whether unit roots are present and where they are located in the system. But because the asymptotic distribution theory changes discontinuously between the stationary and unit root cases, classical hypothesis testing based on asymptotic theory cannot deliver reasonable procedures for inference based on the asymptotic theory.

The central problem with classical inference in these models is that it does not have the usual interpretation as Bayesian inference under some distribution of prior beliefs. This means it will necessarily give unreasonable signals about the relative plausibility of various possible sets of parameter values, as we point out below. However, there are some mechanical difficulties as well, mainly arising out of the discontinuity of the classical asymptotic theory.

Confidence regions based on asymptotic theory will frequently be disconnected because of the discontinuity in the asymptotic theory. This is a familiar fact from existing applications of the asymptotic theory, once it is recognized that whenever a unit root null hypothesis is accepted with the 'correct' asymptotic theory and rejected with the naive theory the confidence region is disconnected. That the standard stationary asymptotic theory rejects the unit root null means that it also rejects some set of nearby parameter settings which imply a stationary model – for which the stationary asymptotic theory is correct. Yet if the point estimate of the model's parameters is inside the stable region of the parameter space, there is also a region of accepted parameter values in the neighborhood of the point estimate. The confidence

region therefore is in two disjoint pieces. This is not just a small sample problem if the true model does indeed contain a unit root. In that case split confidence regions will occur with probability bounded away from 0 as $T \to \infty$.

Consider the case of inference about ρ in the simple univariate AR,

$$y(t) = \rho y(t-1) + u(t), \tag{1}$$

with

$$u(t) \sim \mathrm{N}(0, \sigma^2) \quad \text{independent of past } y,$$

and the parameter space known to exclude true ρ's above 1. If confidence intervals are based on the usual t-statistic (even when $\hat{\rho} > 1$), it is also possible to have two other types of anomaly: empty confidence regions (when $\hat{\rho}$ is too much above 1) and confidence regions not including $\rho = 1$ but including an interval open on the right bounded by $\rho = 1$ (which occurs because the leftward skew of the $\rho = 1$ asymptotic distribution allows easier rejection of $\rho = 1$ when $\hat{\rho} > 1$). In a Monte Carlo study of a univariate autoregressive model with a sample size of 100, $y(0) = 0$, and $\rho = 1$, one of these anomalous cases occurred about 8% of the time in 500 trials, with 3% each in split intervals and empty intervals and 2% in intervals open on the right.

Besides generating confidence regions of disconcerting topology, the discontinuity in the asymptotic theory encourages unreasonable approaches to inference. In an ordinary regression model suppose we entertained as hypotheses two linear restrictions on the coefficients, $R\beta = \gamma$ and $S\beta = \phi$. Suppose also the substantively important restriction were $R\beta = \gamma$ and the other were one we entertained only as a plausible possible simplification of the model. Then, if testing $S\beta = \phi$ rejected even at the 0.15 or 0.10 significance level, it would seem reasonable not to impose $S\beta = \phi$ in testing $R\beta = \gamma$. The data are indicating that $S\beta = \phi$ is unlikely, and we do not want to distort the results of a test of $R\beta = \gamma$ by imposing a restriction the data indicate is probably false. This raises no difficulty because testing $R\beta = \gamma$ in the more general model is still easy. But in a dynamic model, if $S\beta = \phi$ is a restriction implying a unit root and if the distribution of the test statistic for $R\beta = \gamma$ is sensitive to the presence of the unit root, we cannot simply follow the safe course and not impose $S\beta = \phi$. Stock and Watson (1987) apply a sequence of tests for the presence of unit roots preparatory to testing the hypotheses of interest and condition on the 'accepted' pattern of unit roots. This results in inference about important hypotheses hinging on maintained hypotheses about non-stationarity which the data actually indicate to be quite unlikely.

A better classical approach is feasible, using Monte Carlo small sample distribution theory, which will not be discontinuous.[1] One can generate the joint distribution of test statistics of interest for a number of parameter points near the likelihood-maximizing one and compare the likelihood of the observed sample under the various estimated joint distributions. Christiano and Ljungquist (1987) provide an example of analysis of this type. In highly multivariate settings this approach becomes very complicated, however, and doing it systematically eventually leads back to a Bayesian framework – which turns out to be easier!

4. A Bayesian perspective on unit roots econometrics

It has long been recognized that Bayesian inference concerning parameters of linear time series models, conditional on the initial values of the observed sample and Gaussian disturbance distributions, encounters no special difficulties for the case of unit roots. The likelihood, and hence the posterior p.d.f. for a flat prior, is Gaussian in shape regardless of whether or not there are unit (or even explosive) roots. This simple flat-prior Bayesian theory is both a more convenient and a logically sounder starting place for inference than classical hypothesis testing.

How does this theory tell us to deal with, say, the problem of testing Granger non-causality in a system which might possibly have one or more unit roots? It suggests simply using the same F-statistic we would use if unit roots were not present and interpreting the statistic in the same way. What the F-statistic tells us about the likelihood function is the same in both cases. If we are not sure whether nonzero constant or trend terms are present, we can include them in the model and still use the usual F-statistic. Despite the fact that classical inference is sharply affected by whether such terms are present in the unit root case, the Bayesian flat-prior theory is not.

This methodological prescription is so simple that econometricians who have spent much time unraveling the complexities of the classical unit-root asymptotics usually suspect it must contain some hidden fallacy. But the only hidden fallacy is in the classical theory. The flat-prior theory does ignore some potentially important complications which we examine below, but the classical theory also ignores them.

Some find it counterintuitive that an hypothesis which cannot be rejected at, say, the 10% level nontheless may have posterior probability of only, say, 2%.

[1] It appears possible in principle that exact classical small sample theory for likelihood ratio tests could, like the asymptotic theory, lead to disconnected confidence regions. In several 500-trial Monte Carlo studies, however, this was not observed to occur. It may be possible, but I have no proof.

Table 1

Probabilities of test results.

Test result	True disease									
	1	2	3	4	5	6	7	8	9	10
1	0.85	0.075	0	0	0	0	0	0	0	0.013
2	0.021	0.85	0.075	0	0	0	0	0	0	0.014
3	0.020	0.075	0.85	0.075	0	0	0	0	0	0.015
4	0.019	0	0.075	0.85	0.075	0	0	0	0	0.016
5	0.018	0	0	0.075	0.85	0.075	0	0	0	0.017
6	0.017	0	0	0	0.075	0.85	0.075	0	0	0.018
7	0.016	0	0	0	0	0.075	0.85	0.075	0	0.019
8	0.015	0	0	0	0	0	0.075	0.85	0.075	0.020
9	0.014	0	0	0	0	0	0	0.075	0.85	0.021
10	0.013	0	0	0	0	0	0	0	0.075	0.85

Here is a simple example. Suppose a clinician is diagnosing a patient who may have one of 10 equally likely diseases. He receives results from a lab test which has the following characteristics: if the true disease is disease i, the test outcome is 'i' with probability 0.85; if the true disease is i, with $2 \leq i \leq 9$, the test indicates '$i+1$' or '$i-1$' with equal probabilities of 0.075; if the true disease is $i = 10$, the test indicates 'i' with probability $0.01666\dot{} + (i-5)*\varepsilon$, $i = 1,\ldots,9$; and if the true disease is $i = 1$, the test indicates 'i' with probability $0.0166\dot{} + (6-i)*\varepsilon$, $i = 2,\ldots,10$. Table 1 indicates the characteristics of the test with $\varepsilon = 0.001$.

If we observe a test result of 2, a likelihood ratio test for the null hypothesis of $i = 1$ has a marginal p-value of 0.13. That is, $i = 1$ is well within a minimum-length 90% confidence interval. Yet the actual probability that the disease is the first disease, given that the test indicates $i = 2$, is only 0.022. This conclusion does of course depend on the prior – the notion that the diseases are equally likely before we see the test results. But the classical p-value of 0.13 cannot be regarded as simply corresponding to a higher prior weight on the $i = 1$ hypothesis. The marginal p-value for the $i = 1$ null hypothesis if we observe a test result of 10 is below 0.013, yet the probability of $i = 1$ given this observation is 0.014, not greatly different from the 0.022 probability of this hypothesis when the observation is 2. This reflects the actual characteristics of the lab test: if $i = 1$ is the truth, then when the test errs, it errs nearly randomly. An observation of 10 is only slightly stronger evidence against a true $i = 1$ than is an observation of 2. If one puts higher prior weight on the $i = 1$ null hypothesis, then the probabilities of $i = 1$ should be scaled up for all observations in the range 2 to 10. The classical 'p-value' for $i = 1$ changes by a factor of 10 between an observation of 2 and an observation of 10; if the p-value is interpreted as an indicator of the strength of evidence against $i = 1$, it is seriously misleading.

What happens in the foregoing example is less complicated than the situation in an autoregressive model. In the simplest autoregressive model (1) we expect the standard error of estimate of the OLS estimator $\hat{\rho}$ in a sample of given size to be smaller the closer is ρ to 1 (because the sum of squared lagged y's will tend to be larger relative to σ^2 for larger ρ's). This by itself makes it more likely that a given observed $\hat{\rho}$ is a spuriously high estimate generated by a smaller ρ than that it is a spuriously low estimate generated by a larger ρ. But this does not skew the likelihood toward lower ρ's because the distribution of $\hat{\rho}$ is itself skewed to the left for ρ's near 1, which by itself would make it more likely that a given observed $\hat{\rho}$ is spuriously low than that it is spuriously high. The classical theory focuses entirely on this latter effect, paying no attention to the danger that we can be misled into giving too much credence to large ρ values because of the more erratic behavior of estimates from models with lower ρ values.

We can put this same point another way. If the economics profession, perhaps across many investigators and articles, studies many time series which satisfy the simple univariate AR model, and if the models actually have ρ values scattered with some continuous c.d.f. in a neighborhood of $\rho = 1$, then $\rho - \hat{\rho}$, as the error in the Bayesian posterior mean, will be scattered evenly above and below 0. Classical confidence regions, even if constructed from exact finite sample theory, will be skewed toward $\rho = 1$ and will not be centered at the true ρ values on average. No Monte Carlo study is needed to check this proposition; it is simply a characterization in words of what the Bayesian posterior mean is and of how the classical confidence intervals differ from it.

5. Giving unit roots special prior weight

For the reasons outlined in section 2, I think we are mistaken to treat roots exactly on the unit circle as having nonzero probability. We really have no reason for distinguishing such models from their neighbors with roots close to the unit circle. Nonetheless it may be worthwhile to show explicitly how a discrete weight on the unit root hypothesis can be accomodated in a Bayesian framework.

Priors in which there is discrete weight on unit roots are priors with discrete weight concentrated on a lower-dimensional submanifold of the original parameter space. In this respect they are formally similar to models with discrete weight on short-lag-length models embedded in long-lag-length models and to regression models with discrete weight on equations with small numbers of variables embedded in less parsimonious specifications. Practical workers in time series analysis now widely apply the Akaike and Schwartz criteria [see Judge et al. (1980, sect. 11.5)] for choosing lag length because of the obvious unreasonableness of classical hypothesis testing in this problem.

The Schwartz criterion has an asymptotic Bayesian justification. It suggests comparing the twice the difference in log likelihoods between a larger and smaller model (the chi-squared statistic) to the difference in numbers of parameters between the two models, multiplied by log of sample size.

In our simple first-order AR model, the $\rho = 1$ restriction is a zero-dimensional linear subspace of the one-dimensional general parameter space. It might appear that the Schwartz criterion applies directly and suggests comparing

$$\log \frac{\sum (y(t) - y(t-1))^2}{\sum (y(t) - \hat{\rho} y(t-1))^2} \quad \text{to} \quad \log(T),$$

where T is sample size and $\hat{\rho}$ is the least-squares estimate of ρ. This is not correct, however, even asymptotically. The Schwartz criterion's asymptotics depend on the distribution of the estimate converging at the same rate for all true parameter values in a neighborhood of the null hypothesis, and that is not true here.

We can get asymptotically correct procedures, however, along the lines of Leamer's (1978) more general version of the same idea embodied in the Schwarz criterion. Suppose in our univariate AR model we initially put probability α uniformly on the interval $(0,1)$, probability $1-\alpha$ on $\rho = 1$, and independently a uniform prior on $\log(\sigma^2) = \log(\text{var}[u(t)])$. The likelihood then has a normal-inverse-gamma shape, conditional on the initial observations[2] and the marginal likelihood for ρ is a t-distribution with $T-1$ degrees of freedom and scale parameter $\sigma_\rho = \sqrt{\left[\sigma^2 / \sum (y(t-1)^2\right]}$. This distribution is, for large T, very close to $N(\hat{\rho}, \sigma_\rho^2)$. We let Φ be the c.d.f. for the standard Normal distribution and ϕ be its p.d.f. and let $\tau = (1-\hat{\rho})/\sigma_\rho$ stand for the conventional t-statistic for $\rho = 1$. Then for large T the odds ratio in favor of the $\rho = 1$ null hypothesis is $1-\alpha$ times the normally shaped likelihood value at τ divided by α times the integral of the normally shaped likelihood over $(0,1)$. It is easy to see that this is just $(1-\alpha)\phi(\tau)/\sigma_\rho\{\alpha\Phi(\tau)\}$, assuming that the posterior probability on $\rho < 0$ turns out to be negligible. This would reduce asymptotically to the Schwarz criterion were it true that σ_ρ behaves asymptotically like a constant times $1/\sqrt{T}$. Thus the Schwarz criterion would apply for testing $\rho = 0.9$ against a smoothly distributed alternative; but since, when $\rho = 1$, σ_ρ behaves asymptotically like a constant times $1/T$, we cannot avoid including σ_ρ explicitly in the criterion. The analogue of the Schwarz criterion here suggests comparing τ^2 to $-\log(\sigma_\rho^2)$.

[2] That is, it has the form of the product of a gamma marginal p.d.f. on σ^{-2} with a normal conditional p.d.f. for $\rho|\sigma^2$.

Note that the result will not even approximate the results of classical hypothesis testing in general, though it will strongly favor the unit root null hypothesis. The Bayesian procedure gives the unit root hypothesis an edge which depends on σ_ρ. This is reasonable because, if $\rho = 1$ is true, σ_ρ should shrink much faster with sample size than if $\rho \neq 1$. One of the unreasonable aspects of the classical approach to this problem is that likelihood ratio tests make no use of our knowledge that a large σ_ρ in a large sample is evidence against $\rho = 1$ even if the t-statistic for $\rho = 1$ is fairly small.

In any actual sample, the choice of the prior weight α on the open unit interval can have strong effects on the statistic suggested here. In fact, it is probably not reasonable to treat the prior as uniform over all of $(0,1)$. In practice, we are concerned with the situation where the likelihood is concentrated somewhere near 1. The prior is probably also concentrated near there, though we may suppose that with reasonable sample sizes it will nearly always be nearly flat in the range of ρ values for which the likelihood is large. How concentrated the prior will be near 1 obviously depends on the time units in which the data are measured, with concentration getting stronger as time unit gets smaller. If we think of the prior as concentrated mainly on $(0.5,1)$ for annual data, then since $0.5^{1/4} = 0.84$ and $0.5^{1/12} = 0.94$, it should be on $(0.84,1)$ for quarterly data and on $(0.94,1)$ for monthly data. With $\alpha = 0.5$, a t-statistic of 2 for the $\rho = 1$ null hypothesis leaves the odds even between $\rho = 1$ and the alternative if $\hat{\rho} = 0.92$ for the annual data, if $\hat{\rho} = 0.98$ for quarterly data, and if $\hat{\rho} = 0.99$ for monthly data. If α were 0.8 instead of 0.5 (which seems more reasonable to me), then the foregoing $\hat{\rho}$'s at which a t of 2 leaves the odds even correspond instead approximately to the levels at which a t of 1 leaves the odds even.

Though I think it unreasonable to look for exact unit roots, in practice we may need to simplify our models for computational reasons, and then an odds ratio test like that described above is reasonable. To be completely explicit, the criterion would be to compare τ^2 (the squared t-statistic) to

$$2\log((1-\alpha)/\alpha) - \log(\sigma_\rho^2) + 2\log(1 - 2^{-1/S}),$$

where S is the number of periods per year (e.g., 12 for monthly data) and we are ignoring a term $-2\log(\Phi(\tau))$ which will ordinarily be quite small when $\hat{\rho} < 1$ and is asymptotically negligible. Using $\alpha = 0.8$ as I prefer, this becomes for the three leading cases of annual, quarterly and monthly data:

$$-\log(\sigma_\rho^2) - 4.15, \quad -\log(\sigma_\rho^2) - 6.45, \quad -\log(\sigma_\rho^2) - 8.53, \tag{2}$$

respectively.

This approach generalizes to more complicated models, where it amounts to modifying likelihood ratio tests along the lines of Leamer's suggested modifi-

cation of the F-test[3] in single-equation regression models. Of course, in more general models there is usually no one root or pattern of roots on the unit circle which has an a priori special status, and the justification for departing from a continuous prior p.d.f. is even more strained than in the univariate first-order model.

When we are testing a set of linear restrictions, the crude multivariate version of the criterion, corresponding to comparing τ^2 to $-\log(\sigma_\rho^2)$, is to compare the chi-squared statistic (or numerator degrees of freedom times F-statistic) with $-\log|\Sigma|$, where Σ is the covariance matrix of the linear combination of coefficients being restricted. In some cases nonstationarity null hypotheses correspond to sets of linear restrictions – that certain sums of coefficients are 1 or 0. In these cases the reasoning which led to the criteria displayed in (2) generalizes directly, with the criteria in (2) applying separately to each sum of coefficients being tested. For testing k null hypotheses jointly, the correction factors in (2) are each multiplied by k, the statistic becomes the chi-squared instead of the t, and $\log(\sigma_\rho^2)$ is replaced by $\log(|\Sigma|)$.

More generally, though, we might be interested in, say, a single unit root in an m-variable system, so that $m-1$ linear combinations of the variables are stationary. If we don't know a priori which $m-1$ linear combinations are stationary, the generalization of the conventional priors which led to (2) is not so obvious. We will not pursue this point here, simply noting that the analytical difficulties of the further generalization are simpler than those in the classical asymptotic theory for this case.

6. Initial conditions and constant terms

A Bayesian perspective on dynamic models with possible nonstationarity focuses attention on the fact that inference in these models really requires assessment of prior beliefs about whether initial conditions were generated by the same mechanism generating subsequent data. When one is certain the mechanism is stationary and has been operating a long time, then it is reasonable to believe that the parameters of the dynamic model determine the marginal distribution of the initial conditions. In our simplest univariate AR example, this means that $y(0) \sim \mathrm{N}(0, \sigma^2/(1-\rho^2))$. If this were so, inference based on the conditional p.d.f. for $y(t)$, $t=1,\ldots,T$, given $y(0)$ would be throwing away important information. On the other hand, in a nonstationary model there is no unique marginal distribution for $y(0)$, so it does make sense to condition on initial observations.

Our treatment of the stationary and nonstationary cases should not differ discontinuously. A stationary model with a root implying decay toward the mean very slowly relative to the length of the available sample may imply that the model mechanism would have to have been operation for a period many

[3]See Leamer (1978, sect. 4.5).

times longer than the available sample in order to generate convergence of the marginal distribution of the data to its stationary value. It is usually not reasonable to assume the mechanism has operated unchanged for so long, so the stationary marginal distribution should not be taken to be exactly the marginal distribution of $y(0)$. Ideally our prior should use the initial observation in inference for stationary parameter settings and blend smoothly into conditioning on initial observations for nonstationary settings. There are no standard procedures to do this. We may take some comfort in observing that for stationary models the difference between inference conditioned on initial observations and inference taking account of the dependence of $y(0)$ on the parameters will be small in large samples, so that inference conditioned on initial observations may be a good approximation most of the time.

If the model contains constant terms, then it might be natural to take the prior to be flat in the unconditional mean of the data rather than in the constant terms themselves. This idea interacts with the question of whether to condition on initial observations, however, since there is no unique unconditional mean for a nonstationary model.

7. Nonnormality

Bayesian analysis of linear time series models is easy only under the assumption of Gaussian disturbances. The classical asymptotic theory is nonparametric, in that it is unaffected by the form of the distribution of disturbances so long as they are, say, i.i.d. and of finite variance. Though the literature in this area for Bayesian models is much less rich, it is clear that similar nonparametric behavior applies to Bayesian inference. That is, one can assert that with high probability Bayesian inference on normality assumptions will be close to that for the true distribution of disturbances in large samples.[4] Furthermore, in the Bayesian procedure it is clear how to check whether the sample is large enough and how to proceed if it turns out not to be. One can find the peak of the posterior distribution and the second-order derivatives of its logarithm at the peak (this determines the covariance matrix of the Gaussian approximation to it) under normality assumptions, then compare the shape of the approximate Gaussian likelihood to the shape implied by alternative nonnormal disturbance distributions. The Gaussian assumptions allow easy computation of the peak and the second derivative matrix; assessing the accuracy of the approximation requires only evaluating the likelihood under

[4] This is a simplification. The most direct analogue to classical asymptotic results is the result that any true likelihood function converges in shape to a normal distribution. Results like these appear in Hartigan (1983). But this is a guide to convenient computation or summary of the likelihood function, not a robustness result. Robustness results would be of the form: Bayesian inference using the true distribution but conditioned on observation of the Gaussian sufficient statistics (not on the full sample) converges to Bayesian inference using a Gaussian likelihood based on the full sample. I do not know where such results may have appeared, though they are clearly obtainable.

nonnormal assumptions at a number of parameter points, not maximizing and differentiating it. Of course, if the Gaussian approximation turns out to be poor, one can only proceed with the more burdensome computations using the nonnormal likelihood.

The robustness of asymptotic theory to nonnormality does not apply to models with roots strictly larger than 1 in their autoregressive matrices (strictly inside the unit circle in their characteristic polynomials). This applies also to Bayesian inference. Since the data often make roots slightly inside the unit circle look likely, econometricians should be paying more attention to the consequences of nonnormality and to the degree to which the data support a normality hypothesis. My own research (1988) shows strong evidence of nonnormality in macroeconomic time series data. The Bayesian framework makes investigating nonnormality a straightforward computational problem, albeit not an easy one.

8. Conclusion

There are important research problems in time series analysis which interact with the presence of nonstationarity – nonnormality of disturbances and proper accounting for the evidence about parameters contained in initial conditions, for example. A Bayesian framework for inference focuses attention on these problems and suggests approaches to dealing with them. Classical inferential procedures, precisely because they do differ substantially from Bayesian procedures in this context, are misleading, and they throw up analytical difficulties in simple cases which prevent our making progress on the real issues.

References

Christiano, L. and L. Ljungquist, 1987, Money does Granger-cause output in the bivariate output–money relation (Federal Reserve Bank of Minneapolis, MN) in process.

Granger, C., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424–438.

Hall, R.E., 1978, Stochastic implications of the life cycle permanent income hypothesis: Theory and evidence, Journal of Political Economy 86, 971–988.

Hartigan, J.A., 1983, Bayes theory (Springer-Verlag, New York).

Judge, G.E., W.E. Griffiths, R.C. Hill and T.-C. Lee, 1980, The theory and practice of econometrics (Wiley, New York).

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

Lucas, R.E., Jr., 1978, Asset prices in an exchange economy, Econometrica 46, 1429–1446.

Sims, C., 1984, Martingale-like behavior of prices and interest rates, Center for Economic Research discussion paper no. 205 (University of Minnesota, Minneapolis, MN).

Sims, C., 1988, A nine variable probabilistic macroeconomic forecasting model (Federal Reserve Bank of Minneapolis, MN) in process.

Sims, C., J. Stock and M. Watson, 1986, Inference in autoregressive models with some unit roots, Discussion paper available from the authors.

Stock, J. and M. Watson, 1987, Interpreting the evidence on money–income causality (Hoover Institution, Washington, DC) in process.