On DeJong and Whiteman’s Bayesian inference for the unit root model

Fallaw Sowell*
Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received April 1990, final version received May 1991

For a Bayesian approach to be useful, the priors and posteriors must be carefully interpreted to guarantee that the economic conclusions reached are justified; otherwise, the approach simply yields a set of summary statistics with little economic insight. The Bayesian approach to testing between different models of trend proposed in DeJong and Whiteman (1991) suffers from this shortcoming.

1. Introduction

A dominate issue during the past fifteen years has been the appropriate statistical model for the trend in macroeconomic time series. The importance of this issue rests on the facts that different models of the trend can imply different conclusions concerning the validity of economic theories and can also imply different policy implications for macroeconomic models. To date, most of the econometric research concerning trend behavior has followed a classical statistical approach. These test statistics have been applied in numerous empirical studies, often with imprecise results. For example, it is not uncommon for a series to be consistent with both the null hypothesis of a unit root model and also the null hypothesis of a time trend model. Several researchers have noted the cause of this situation is the low power that these classical test statistics have to plausible alternative models. Currently there exists a large number of (consistent) tests, but given the current size of our data sets it is often difficult to distinguish between reasonable models of trend.

*I would like to thank Jay Kadane, Charles Plosser, Charles Whiteman, and Stan Zin for helpful conversations. Financial support by Carnegie Mellon University through a faculty development grant is acknowledged.
It has been suggested that by focusing the data through a Bayesian framework, more precise statements can be made concerning the appropriate trend behavior of macroeconomic time series. Rather than making statements concerning the consistency of the data with a given null hypothesis, a Bayesian approach can make statements concerning the relative support the data has for different hypotheses. Such an approach could prove useful by allowing economists to focus attention on those models that most likely generated the data. However, for a Bayesian approach to be useful the priors and the posteriors must be carefully interpreted to guarantee that the economic conclusions reached are justified; otherwise, the approach simply yields a set of summary statistics with little economic insight. The Bayesian approach to testing between different models of trend proposed in DeJong and Whiteman (1991) (referred to as DJW) suffers from this shortcoming.

In the following sections, problems with the approach used in DJW are noted. The general conclusion is that nothing is learned of the relative support in the data for the unit root and the time trend models because the unit root model was never formally considered in the analysis. Bayesian techniques may give economists new insights into the question of the trend behavior of macroeconomic time series, but only if the relevant models of trend are actually considered.

The question DJW claims to address is the support that various macroeconomics series have for the hypothesis that many macroeconomic time series appear to be consistent with a unit root model. Unfortunately, the techniques used do not directly address this question. This does not mean that the procedures used in DJW were performed incorrectly. The issues concern the appropriateness of the technique, the interpretation of the posteriors, and the generality of the conclusions. The next section discusses general problems with the arguments and methodology used in DJW. The third section is concerned with the interpretation of the densities reported in DJW. The fourth section considers the procedures presented in DJW as a classical hypothesis testing procedure. The final section contains some concluding thoughts.

2. The methodology of DJW

The first problem with DJW concerns the inconsistency between the approach and the conclusions. DJW demonstrates priors that are sufficient for various posterior distributions, but makes unqualified statements that suggest necessary conditions for the posteriors. In DJW, analysis is restricted to AR(3) models and certain interpretations are given. These interpretations are then claimed to hold in general (emphasis added)

'We employ a Bayesian perspective to identify the type of prior needed to support the inference that most macroeconomic time series follow ran-
dom walks. For many of the series considered by Nelson and Ploesser (1982) the required prior involves assigning very low probability to trend-stationary alternatives.

These general statements made in the abstract of DJW should not be claimed. Only a small set of priors are considered in DJW. The interpretations can only be claimed to hold conditional on restricting attention to the priors considered. Because this set is not exhaustive, the conclusions cannot be claimed in general.

The importance of this argument is underscored by the fact that different Bayesian conclusions are possible if different priors had been used. This has been explicitly demonstrated in a recent paper that used an alternative prior. In Phillips (1990) an alternative ignorance prior is considered and different posterior densities are achieved. Researchers cannot embrace the interpretations presented in DJW without first embracing their priors and disregard all other possible priors.

A second problem with DJW is the prior probability assigned the unit root model. The question supposedly considered in DJW concerns the data's relative support for the unit root model. However, this model is always assigned zero prior probability. Hence there is always zero posterior probability that the data follow a unit root model. In DJW this inconsistency between the question considered in the paper and the technique used is resolved by an arbitrary rule to 'infer that the unit root is implausible if there is not at least 5% posterior probability that the root exceeds 0.975'. There is no formal justification for this rule.

Standard Bayesian inference concerning a point null hypothesis, such as the unit root hypothesis, would place some finite prior probability on the null hypothesis. One reason for not considering mass at the point null hypothesis are the arguments presented in Sims (1988). The arguments raised in Sims (1988) are economic arguments that suggest economists have incorrectly given special attention to the unit root model. The implication of these economic arguments would be that economists should not be concerned with the unit root hypothesis. Of course, the primary issue in DJW is the unit root hypothesis. Sims' economic arguments should not be confused with a statistical argument concerning the appropriateness of placing prior mass at a single point in the parameter space. If Bayesian inference is being conducted on a point null hypothesis, then it should be done correctly, with mass on the hypothesis.

3. Interpretation of the posteriors and the priors

DJW considers two prior distributions for the model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \delta t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2),$$

(1)
for macroeconomic time series. The first prior will be denoted
\[ \pi_A(\theta) = p(\theta) \iota_A(\theta), \]
where \( \theta = (\beta_0, \beta_1, \beta_2, \beta_3, \delta, \sigma) \), \( p(\theta) = 1/\sigma \) is an improper diffuse prior, and \( \iota_A(\theta) \) is an indicator function over the set \( A = \{(\delta, \Lambda): 0 \leq \delta < 0.016, 0.55 \leq \Lambda < 1.055\} \) (recall that \( \Lambda \) is the modulus of the dominate AR root). The second prior will be denoted
\[ \pi_B(\theta) = p(\theta) \iota_B(\theta), \]
where \( \iota_B(\theta) \) is an indicator function over the set \( B = \{(\delta, \Lambda): \delta = 0.0, 0.65 \leq \Lambda < 1.155\} \).

DJW reports two general results
(I) Conditional on \( \pi_B(\theta) \) the mass of \( \Lambda \) is centered around one.
(II) Conditional on \( \pi_A(\theta) \) the mass of \( \Lambda \) is centered below one.

The analysis that derived these general results is correct; however, some of the interpretation that DJW gives these results are flawed. In DJW it is argued that result (I) is support for a unit root model and is the Bayesian analog of the results reported in Nelson and Plosser (1982); but this is not the case. The Nelson and Plosser (1982) regressions contain a time trend, but the regressions that established result (I) did not. The dependent variable in Nelson and Plosser (1982) was the first difference of the series, not the level used in DJW. Furthermore, the claim from Nelson and Plosser (1982) was that the data were consistent with a unit root model, while the posterior probability referred to in result (I) assigns zero probability to the unit root model. For all the priors and posteriors considered in DJW, the mass associated with \( \Lambda < 1 \) supports a stationary model and the mass associated with \( \Lambda > 1 \) supports an explosive nonstationary model. Formally neither the priors nor the posteriors show support for a unit root model. The classical results reported in Nelson and Plosser (1982) are not comparable to any of the Bayesian results reported in DJW.

The decision rule that DJW gives to judge support for a unit root model is also incorrect. As noted above, DJW uses the arbitrary rule that posterior mass associated with values of \( 0.975 < \Lambda \) is support for a unit root hypothesis. This rule is inappropriate. Mass associated with \( 0.975 < \Lambda < 1.0 \) is support for a stationary model. Mass associated with \( 1.0 < \Lambda \) is support for an explosive nonstationary model. Only mass associated with \( \Lambda = 1.0 \) is support for a unit root model. The decision rule used in DJW to judge support for a unit root model is apparently based on some idea of how the posteriors will

---

1Both priors considered in DJW have zero mass associated with the unit root model. By these priors, the data are restricted in terms of how they can be informative. Because the prior is continuous in \( \Lambda \) the data can only shift mass along the line. It does not have any possibility of showing point mass on the point \( \Lambda = 1 \).
behave under a misspecified model. Unfortunately, there is no theory for Bayesian analysis when the prior has been misspecified. There is no guidance for interpreting these posterior densities when the prior densities assign zero probability to the model of interest.

In Bayesian analysis the results are conditional on the priors, hence the accurate statement of a researcher's prior beliefs is a critical step in the analysis. It is possible for researchers with different priors to analyze the same data and arrive at different conclusions. To help the readers judge the appropriateness of the priors used in DJW some of the implications are listed.

1. There is zero probability of a unit root model.
   (a) The prior \( \pi_A(\theta) \) has support in \( \mathbb{R}^5 \), but the hypothesis of a unit root is defined in a subspace of \( \mathbb{R}^3 \) (\( \delta = 0.0 \) and \( \Lambda = 1.0 \)).
   (b) The prior \( \pi_B(\theta) \) has support in \( \mathbb{R}^4 \) and the hypothesis of a unit root is defined in a subspace of \( \mathbb{R}^3 \) (\( \Lambda = 1.0 \)).

2. For the prior \( \pi_A(\theta) \) the probability that there is a time trend is one.

3. For both \( \pi_A(\theta) \) and \( \pi_B(\theta) \) there is zero probability of the correct model being an AR(\( p \)) with \( p \neq 3 \).\(^2\)

4. The data are generated from a Gaussian distribution, with probability one.

5. For both \( \pi_A(\theta) \) and \( \pi_B(\theta) \) there is zero probability of there being MA terms.

6. For both \( \pi_A(\theta) \) and \( \pi_B(\theta) \) there is finite probability that the largest root of the AR polynomial is explosive. In DJW it is argued that this 'favors difference-stationarity over trend-stationarity'. Just the opposite is true. Zero mass is assigned to \( \Lambda = 1 \), which implies zero support for difference-stationarity. Stationary roots and explosive roots are both consistent with a time trend.

7. For both \( \pi_A(\theta) \) and \( \pi_B(\theta) \) there is prior probability that all three roots of the AR polynomial are explosive.

With the strong implications of both \( \pi_A(\theta) \) and \( \pi_B(\theta) \) in mind, it is important to recall that Bayesian results may not be robust to misspecification. The accurate specification of the model (i.e., the prior) is crucial. Without knowing the robustness of the results to the priors it is inappropriate for another researcher to personally embrace the conclusions of DJW unless the researcher's beliefs were those used in DJW. In the unit root literature this has often not been an issue. The reason is that several classical unit root

\(^2\)It is inappropriate to use a classical \( F \)-test to determine the length of the AR polynomial, as in DJW. A researcher's priors must be independent of the data. Priors should be determined by theory and data which are independent of the data to be analyzed.
tests are based on the invariance principle which implies asymptotic distributions are robust to some misspecification. However, this is not the case for a Bayesian analysis.

So what do the posteriors reported in DJW show? Results (I) and (II) show that when a time trend is included, then the data are more consistent with a stationary model rather than an explosive model. The data cannot be informative concerning the validity of the unit root because there is always zero prior probability that the models follow a unit root model. The posteriors do not give any direct information concerning relative support for a unit root model.

4. A classical hypothesis test?

The majority of econometricians are trained as classical statistician. How should these researchers interpret the results in DJW? Can the procedure in DJW be used to distinguish between unit root and time trend models? Regardless of the paradigm under which the procedure was developed, if the unit root and the time trend models can be distinguished by this procedure then it should be used extensively. Unfortunately, the procedure cannot. As demonstrated below, the posteriors reported are consistent with both models.

For every series, the decision rule used in DJW was based on the statistic 'the posterior probability that the modulus of the dominate root is greater than or equal to 0.975'. This statistic will be referred to hereafter as simply \( \text{Pr}(\Lambda \geq 0.975) \). From a classical statistics point of view, this probability can be calculated for any sample and can be thought of as a test statistic. However, to interpret the observed value for a given sample, a classical statistician would be concerned with the sampling distribution of this statistic. For example, if the data follows a given unit root model, how likely is an observed value of the statistic? The inability of this statistic to distinguish reasonable alternative models can be demonstrated by calculating the sampling distributions of this statistic for alternative models.

The following experiment is reported below. Consider two AR(3) models for the real GNP series used in DJW. Let model I include a drift term and let model II include both a drift and a time trend. For each model simulate random samples. For each simulated sample calculate \( \text{Pr}(\Lambda \geq 0.975) \) and use these values to estimate the sampling distribution of the statistic under the two different models.

The parameters of the 'reasonable alternative' models will be the OLS estimates for the real GNP series. Model I is actually a unit root model,

\[
y_t = 0.03 + 1.37y_{t-1} - 0.45y_{t-2} + 0.08y_{t-3} + \varepsilon_t,
\]

\[
\varepsilon_t \sim \text{iid } \mathcal{N}(0.0040),
\]
and model II is a time trend model,

\[ y_t = 0.89 + 1.22y_{t-1} - 0.35y_{t-2} - 0.05y_{t-3} + 0.006t + \epsilon_t, \]

\[ \epsilon_t \sim \text{iid } N(0.0035). \]

For both models 500 samples of length 62 were generated. For each sample the first three observations were fixed at the initial values of the series considered in DJW. The procedure outlined in DJW was performed for each series with 20000 draws from each posterior. This gave 500 draws from the distribution of the statistic \( \Pr(\lambda \geq 0.975) \) for each model. The density for the 500 values estimates the sampling density for the statistic under each model and are plotted in fig. 1. Fig. 1 shows that many values of the statistic (e.g., \( \Pr(\lambda \geq 0.975) \in [0.05] \)) are consistent with both models. This highlights the difficulty that this statistic has in distinguishing these two models for real GNP.

The test statistic value for the observed real GNP series was estimated to be\(^3 0.00837; \) this was estimated using 400000 draws from the posterior so that

\(^3\)The value reported in DJW is 0.003; however, the numerical standard error associated with this estimate is larger than 0.0035. Note that one standard error away from the point estimate includes negative values. Hence a more accurate estimate was derived. The point estimate reported here is easily within two standard errors of the point estimate reported in DJW.
the numerical standard error is below 0.0008. For the two possible models, how surprising is it that we observed \( \Pr(A > 0.975) = 0.00837 \)? For model II, the time trend model, 218 of the 500 statistics were greater than 0.00837 giving an empirical \( p \)-value of 44.4\%, which as noted by DJW shows that the observed real GNP series is consistent with a time trend model. However, for model I 76 of the 500 statistics were less than the observed value giving an empirical \( p \)-value of 15.2\%. The observed statistic is also consistent with a unit root model and cannot be ruled out based on the observed statistic. Three out of twenty times when the series is generated from the unit root model given by model I the observed posterior probability that \( \{A \geq 0.975\} \) is less than 0.00837. This demonstrates that (for real GNP) the reported results are consistent with both models.

Of course for a Bayesian statistician this issue is irrelevant. There is only one realization of real GNP from 1909 to 1970, so questions concerning: 'What if a different series had occurred?' are not important. But this does underscore the importance of not interpreting the Bayesian results reported in DJW as classical statistics.

The experiment above does have an important implication for Bayesian statisticians concerning the interpretation of the Bayesian results in DJW. To decide if the data support a unit root model, DJW report the posterior probability that \( \{A \geq 0.975\} \). The apparent intuition is that if the series really follows a random walk, then mass must be centered above 0.975. However, this intuition is incorrect, as is demonstrated in fig. 1 by the density for the test statistic when the series follows the unit root model of Model I. Caution must be exercised when interpreting posteriors derived from Bayesian analysis concerning misspecified models.

5. Conclusions

DJW accurately reports the posterior densities conditional on the data and their priors. However, these posteriors should not be viewed as supporting a deterministic trend model over a unit root model. The basic problem is that the unit root model was never formally considered in the analysis presented in DJW.

The question of the correct model for the trend behavior for macroeconomic time series remains unanswered. It is still unclear if a univariate Bayesian approach can give new insights into this question. The problem is not inherent in the statistical approach, but rather the lack of data. This is also a problem with classical tests. There may not be enough long-run information in currently available macroeconomic series to distinguish alternative models of trend with univariate tests. Several strategies remain available to address this question. One possibility is for economic theory to give additional structure to the models. Additional structure may restrict the class
of models that the data needs to distinguish. An alternative approach would be to focus the long run-information in several series by estimating multivariate models. The additional precision associated with multivariate models may be sufficient to distinguish alternative models. As long as economic theories and policy implication are sensitive to the models of trend, economists must be concerned with distinguishing these models.

References


