

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 447

July 1993

**NEAR OBSERVATIONAL EQUIVALENCE AND UNIT ROOT
PROCESSES: FORMAL CONCEPTS AND IMPLICATIONS**

Jon Faust

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Abstract

A number of recent papers have discussed the fact that difference stationary and trend stationary processes are *nearly observationally equivalent*. The meaning of this fact, however, remains clouded. This paper defines near observational equivalence and derives several implications of the notion for classical and Bayesian unit root inference. For example, unless restrictions are imposed on the general difference and trend stationary models, the exact size of any consistent unit root test rises to one with sample size. Bayesian posteriors regarding unit roots are arbitrary in the sense that given any prior, there are other priors that agree with the first regarding empirical outcomes, but that imply arbitrarily different unit root posteriors.

Near observational equivalence and unit root processes: formal concepts and implications

Jon Faust¹

Processes with and without unit roots are nearly observationally equivalent. Statements such as this are becoming common in discussions of unit root issues [Blough 1989,1991; Christiano and Eichenbaum 1990; Campbell and Perron, 1991; Cochrane 1991], but the precise meaning of such statements remain clouded. As a result, a number of potentially misleading claims and conjectures have characterized discussion in this area. This paper makes precise several senses in which data are uninformative about the existence of unit roots.

The paper formalizes a concept of *near observational equivalence* that carries the practical implications of strict observational equivalence. This concept provides the basis for an exploration of problems with classical and Bayesian analysis of unit roots. The results flow from a single proposition that can be viewed as a reinterpretation of an important result of Christopher Sims's [1971,1972]. Sims's result is stated in terms of the continuity of loss functions under certain topologies on the parameter space. Re-stating the result in terms of the familiar notion of observational equivalence allows one easily to trace out implications that have been the subject of considerable speculation.

Many of the paper's results are stated in terms of the long-run effect of shocks to a variable, rather than in terms of autoregressive roots of the variable's time series process. A well-known fact, made precise below, is that processes in which shocks have no long-run effect are trend stationary (TS), whereas shocks have a permanent effect in difference stationary (DS) or unit root processes.

Under very general assumptions several facts are shown. First, for every DS process there are nearly observationally equivalent TS processes. For every TS

¹ The author is a staff economist at the International Finance Division of the Board of Governors of the Federal Reserve System. The author thanks Steve Blough, Neil Ericsson, Pierre Perron, Harald Uhlig, and especially Dave Gordon and Christian Gilles for useful discussions.

process there are nearly observationally equivalent DS processes, which may have arbitrarily large long-run effect of shocks. This fact resolves questions about whether DS processes that look TS must have small long-run effects.² The result is not merely a curiosity. Near observational equivalence would not be very troubling if it only implied difficulty in distinguishing a process with a zero long-run effect of shocks from ones in which the effect were tiny. This result says we cannot distinguish any values for the long-run effect of shocks.

Near observational equivalence has two clear implications for classical testing. First, classical tests to distinguish general TS and DS models, no matter which is the null, have power less than or equal to size. Second, if a unit root test is consistent, its size converges to one with sample size. While Blough [1991] and Campbell and Perron [1991] reported the first result in weaker form, the second result is new. It implies that the exact size of most unit root tests run in practice converges to one with sample size—if the tests are viewed as tests of general TS vs. DS hypotheses.

Partly in response to problems with classical unit root testing, Bayesian approaches have received a great deal of attention [e.g., Sims, 1988; Sims and Uhlig, 1991; DeJong and Whiteman, 1993]. This work has precipitated a lively controversy.³ Much attention in that debate has focussed on the fragility of results to specification of the prior. Below, I sidestep all of these issues and show a more generic fragility. Given any prior over the DS and TS parameter spaces, there is another prior with the same outlook regarding observable implications, but disagreeing to an arbitrary degree about the long-run effect of shocks. Thus, even if you agree that the prior reported in some empirical work correctly captures your feelings about various outcomes, you need not have any interest in the posterior over unit root issues. There will be another prior treating outcomes in the same manner, but for which the posterior is arbitrarily different.

The results flow from a fundamental property of DS and TS processes laid out

² Cochrane [1991] conjectures that such DS processes must have small long-run effects; Campbell and Perron [1991] make the analogous claim in the multivariate context.

³ See the debate initiated by Phillips [1991] and the following comments by Sims and others.

in Section 2. Sections 3 and 4 explore the classical and Bayesian implications, respectively; Section 5 extends the results to the multivariate case; and Section 6 concludes.

1 A fundamental property of DS and TS processes

All of the results below are for a general difference stationary model. After laying out this model and the single restriction that trend stationarity places on the model, the section derives a general convergence property for DS and TS structures.

Any realization, $Y^T = (Y_1, \dots, Y_T)$, from a discrete univariate time series process can be represented,

$$Y_t = Y_0 + \sum_{i=1}^t (1 - L)Y_i \quad (1)$$

$t = 1, \dots, T$. Three assumptions define the model of interest here. If the model is to be called difference stationary, we need,

A 1 *The first difference of Y_t , $t = 1, 2, \dots, T$, is covariance stationary.*

Thus, Y_t is at most integrated of order 1 and $(1 - L)Y_t$ has a Wold representation,

$$(1 - L)Y_t = A(L)\nu_t + d_t \quad (2)$$

where d_t ($n \times 1$) is linearly deterministic; ν_t , has mean zero, constant, finite variance and is not serially correlated; $A(L) = \sum_{i=0}^{\infty} a_i L^i$; and $\sum_{i=0}^{\infty} a_i^2 < \infty$.

Under A1, the model is described by $A(L)$, the processes for ν and d , and the distribution of Y_0 . The crucial feature of the model for our purposes is $A(L)$. The remaining elements, Y_0 , ν and d , are parameterized by $\psi \in \Psi$.

A 2 *There are no constraints between the parameters of ψ and $A(L)$.*

The restriction between $A(L)$ and the parameters of Y_0 may seem questionable. We could just as well assume, following most work on the subject, that Y_0 is fixed, or we could push arbitrarily far into the past the date at which the distribution of some Y_t is unrelated to the distribution of sample shocks. This assumption

is generally appropriate, since we have little basis for views about how economic variables evolved in the distant, pre-sample past.

A 3 *The Wold representation of $(1 - L)Y_t$, (2), satisfies $\sum_{i=0}^{\infty} |a_i| < \infty$.*

This assumption is convenient and conventional, allowing us to talk unambiguously about $A(1) = \sum_{i=0}^{\infty} a_i$, which is the long-run effect of each shock, ν_t , to Y .⁴

The sole restriction trend stationarity places on the general model is $A(1) = 0$: the long-run effect of shocks to a trend stationary process is zero. To see this, note that if Y_t^* is (trend) stationary, it can be written $Y_t^* = A^*(L)\nu_t + d_t^*$, with $A^*(1)$ finite.⁵ Differencing Y_t^* to put it in the form of (2) above gives,

$$(1 - L)Y_t^* = (1 - L)A^*(L)\nu_t + (1 - L)d_t^*$$

If d_t^* is linearly deterministic then its first difference is as well. Thus, the only restriction of the trend stationary model is that $A(L)$ can be factored $A(L) = (1 - L)A^*(L)$, where $A^*(1)$ is finite, implying $A(1) = 0$.

The model is described by $A(L)$ and ψ . Under A3, to each $A(L)$ we can associate an infinite sequence $A = \{a_1, a_2, \dots\} \in l^1$, where l^1 is the space of absolutely summable sequences ($a_0 = 1$ by convention). I will refer to A as the moving average (MA) representation of a process with lag polynomial $A(L)$. Thus, the parameters of the model are given by $\theta = \{A, \psi\}$, $\theta \in \Theta$, where under A2, $\Theta = l^1 \times \Psi$. Given this parameterization of the model, for each sample size T , and each $\theta \in \Theta$ there will correspond a distribution function $F^T(\cdot|\theta)$ and an associated probability measure on the Borel sets of R^T , μ_{θ}^T . The T superscripts may be dropped where no ambiguity should result.

⁴ The processes with square, but not absolutely, summable polynomials are called *long memory processes*. Many of the results of the paper can be extended to these processes.

⁵ The deterministic part of the process does not enter the discussions below and is largely ignored. However, if d_t^* includes a time trend, then Y_t^* is called *trend stationary*, rather than simply *stationary*.

1.1 A general convergence property of the DS model

The driving result of the paper is that it is difficult to distinguish two processes whose MA representations, A and A' , are close in the sense that $(\sum_{i=0}^{\infty} (a_i - a'_i)^2)^{1/2}$ is small. Loosely speaking, if the MA representations of two variables are close in the l^2 norm then the distributions of the random variables are similar.

Proposition 1 *Fix any $T > 0$ and $\psi \in \Psi$. Given an MA representation, A , and a sequence of MA representations, $\{A_k\}$, if $\|A_k - A\|_2 \rightarrow 0$, then*

$$F(\cdot|A_k, \psi) \Rightarrow_k F(\cdot|A, \psi)$$

where $\|\cdot\|_2$ is the l^2 norm, and \Rightarrow means convergence in distribution of the underlying random variables.

Proof: See the Appendix.

This is basically a re-statement of the central result in Sims [1972]. Sims showed that a mean-square-error-based loss function is topologically equivalent to the l^2 norm, and the same logic is used in the proof of this proposition. Consider why the first observation, Y_{1k} , converges in distribution to Y_1 . By definition,

$$Y_1 - Y_{1k} = \sum_{i=1}^{\infty} (a_i - a_{ik}) \nu_{1-i}$$

The mean square error from approximating Y_1 by Y_{1k} is

$$E(|Y_{1k} - Y_1|^2) = \sigma^2 \sum_{i=1}^{\infty} (a_{ik} - a_i)^2 = \sigma^2 \|A - A_k\|_2^2$$

where σ^2 is the variance of the ν process. Convergence in mean square, and hence, in distribution follows by noting that expectation goes to zero with k whenever $\|A - A_k\|_2^2$ does.

The following general fact about square summable sequences makes the prospects for learning about $A(1)$ dim indeed: closeness of two MA representations in the l^2 norm has *precisely no implications* for closeness of the sums of coefficients of the representation. As an illustration, take A , that sums to a . Construct A' by adding $(a' - a)/k$ to the first k elements of A . The sequence A' sums to an arbitrary a' . The l^2 distance between A and A' is $(a - a')/\sqrt{k}$, which can be made arbitrarily

small by choosing k large. Formally, for every scalar a , the set of sequences with sum a is dense in l^1 under the l^2 metric.

1.2 Some intuition

Before proceeding to the results that flow from prop. 1, it may be useful to provide some intuition about the processes giving rise to near observational equivalence. In particular, one might wonder whether DS processes that have large long-run effect of shocks and yet behave like TS processes look *strange* in some sense. In standard time domain representations, the answer is a clear no: the DS processes look like the TS processes. Viewing A and A' from above as MA representations, A' differs from A only in that the first k coefficients have been altered by a tiny amount (for large k). Similarly, for large k , the autocorrelation functions for the two processes will look almost identical.

Of course, there are transformations of A and A' that highlight their differences. In this case the frequency domain is illuminating. The spectral density at frequency zero of the $(1 - L)Y_t$ is proportional to $A(1)^2$. The intuition for how a DS process with a large frequency zero spectrum can have similar empirical implications to a TS process with a zero spectrum at frequency zero, is easy to see: the DS process must have similar spectrum to the TS process at all but the lowest frequencies, and then turn upward near frequency zero. The similarity of the spectra at high frequencies makes the two processes difficult to distinguish in finite data.

2 Near observational equivalence and classical inference

In this section, I propose a definition of near observational equivalence that preserves many of the implications of the strict notion. Several implications of near observational equivalence for the TS/DS case are given.

The traditional notion of observational equivalence states that structures param-

eterized by θ_q and θ_r are observationally equivalent if and only if $F(\gamma|\theta_q) = F(\gamma|\theta_r)$ for all γ . Any observation generated by θ_q might just as well have been generated by θ_r . It seems reasonable to say some structure q is *nearly observationally equivalent* to r if $F(\cdot|\theta_q)$ is close, in some sense, to $F(\cdot|\theta_r)$. One formalization of this notion is:

Definition 1 *The sequence of structures parameterized by $\{\theta_k\}$ has nearly observationally equivalent members to a structure θ_r if for any fixed T , $F(\cdot|\theta_k) \Rightarrow_k F(\cdot|\theta_r)$.*

In many senses, this represents only a slight loosening of strict observational equivalence. For example, strict equivalence of θ_q and θ_r implies $F(\gamma|\theta_q) - F(\gamma|\theta_r) = 0$ for all γ . With continuous distributions, near equivalence of members of the sequence $\{\theta_k\}$ to θ_r implies that for any $\varepsilon > 0$ we can find a k such that $F(\gamma|\theta_k) - F(\gamma|\theta_r) < \varepsilon$ for all γ .⁶

2.1 Implications of near observational equivalence

Based on the definition of near observational equivalence and prop. 1, it is clear that,

Proposition 2 *i) To every TS structure there are nearly observationally equivalent DS structures with arbitrary long-run effect of shocks. ii) To every DS structure, there are nearly observationally equivalent TS structures.*

Proof: The proof follows directly from prop. (1) and the denseness of sequences with any sum in l^1 .

This result says that almost no empirical issue turns on the value of $A(1)$: inference about unit roots, in general, makes no more sense than inference regarding strictly unidentified parameters. To formalize this claim, define the TS and DS hypotheses:

$$H_{TS} : A(1) = 0$$

$$H_{DS} : |A(1)| > \eta \geq 0$$

I allow the possibility of $\eta > 0$, to emphasize that the problems discussed here do not stem from difficulty in distinguishing DS structures with tiny $A(1)$ from

⁶ Convergence in distribution with continuous distribution functions is uniform, see e.g., Billingsly, 1979.

TS structures. A central result regarding testing is most easily stated by limiting discussion to probability distributions that can be represented by a density.⁷

A 4 *For all θ and T , $\mu_\theta(\cdot)$ is absolutely continuous with respect to the Lebesgue measure.*

Under this assumption, the following result follows from prop. 2:

Proposition 3 *Assuming A4, consider tests of H_{TS} vs. H_{DS} , regardless of which is the null hypothesis: i) Power is less than or equal to size. ii) Suppose that under some test procedure the rejection probability converges to one with sample size when some structure consistent with the alternative hypothesis is true. The exact size of the procedure converges to one with sample size.*

Proof: See the Appendix.

Since the intuition behind both parts is the same, consider part (ii). Pick some alternative hypothesis structure, θ_r , for which the test is consistent. In any sample size, we can find a null structure that is rejected just as often as θ_r . Since this is true in any sample size, as the sample size grows, and the rejection probability for θ_r increases to one, we are able to find null structures whose rejection probabilities are similarly close to one.

I paraphrase this result by saying that if a test is consistent, then its size goes to one. Others, however, may reserve the word *consistent* for tests of fixed size. In any case, the result applies to most unit root test procedures used in practice, *when they are viewed as tests of the general TS vs. DS hypotheses*. Most test procedures are specified to attain some fixed nominal size—the size suggested by some asymptotic approximation.⁸ The procedures are invariably designed so that the probability of rejection converges to one with sample size for some (or all) structures consistent with the alternative hypothesis. Result (ii) says that the exact size of all these test procedures converges to one with sample size. This result provides a general

⁷ This assumption is needed to rule out test procedures with mass on the boundary of the rejection region.

⁸ The behavior of exact size under result (ii) is consistent with the nominal size of the test being fixed based on some asymptotic distribution. These asymptotics provide only pointwise convergence to the asymptotic distribution for each fixed null structure. Because this convergence is not uniform, it is always possible, loosely speaking, to find a null structure for which convergence has not yet begun.

umbrella accounting for a number of analytic and Monte Carlo results regarding the size distortion in tests of the DS and TS hypotheses [e.g., Schwert, 1988; Hall, 1988; Campbell and Perron 1991; Blough 1989].

2.2 A fix-up for the classical statistician?

The problem with unit root tests is that absent far stronger restrictions on $A(L)$ than those imposed by the general *DS* and *TS* models, finite data support no inference about $A(1)$. As always in cases of observational equivalence, however, imposition of restrictions can identify parameters of interest. Based on this prospect, some authors [Cochrane, 1991; Campbell and Perron, 1991] have suggested that we could define finite-dimensional versions of the hypotheses that rule out certain *problematic* structures, thereby avoiding the near observational equivalence problem. Further, we might append T subscripts to the hypotheses and allow the range of structures covered by the restricted hypotheses to grow slowly with sample size. Done cleverly, this could avoid near observational equivalence at each sample size, but allow the restricted hypotheses to asymptotically approach the hypotheses of interest.

Sims [1971] demonstrated several results for the growing hypothesis approach.⁹ I see the results primarily as demonstrating that the approach is of little practical interest in the case at hand. For example, one can define a sequence of growing, finite-dimensional hypotheses that will allow asymptotically valid inference regarding $A(1)$ —*conditional on one of the two hypotheses being true for some finite T* .¹⁰ For this approach to be of interest, one must minimally be willing to assume that the growing finite-dimensional hypotheses will ultimately contain the true parameter. This amounts to restricting A to lie in a small subset of l^1 .¹¹

Even if one is willing to rule out much of the parameter space of the general

⁹ Berk [1974] shows how to form consistent, asymptotically normal estimators, but not how to perform valid inference about unit root parameters. Inference regarding parameters estimated using Berk's approach are subject to the problems of prop. 3.

¹⁰ Alternatively, one can perform asymptotically valid inference regarding a projection of the true A onto the parameter set consistent with the restricted hypotheses in the sample size of interest.

¹¹ A small subset here is a *meager* or *first category* subset—a countable union of nowhere dense sets, see Oxtoby [1980].

model, this approach is of little comfort. There will be multiple ways to select restricted hypotheses. In any finite sample, the answer one gets in under the restricted hypotheses approach depends critically on the particular restriction chosen. In my view,¹² there is generally little basis in economics for choosing one restriction over another. Thus, this approach requires making an essentially arbitrary choice of one small portion of the parameter space as opposed to another. This arbitrary restriction leads to similarly arbitrary test results. The next section puts a finer point on the manner in which the choice of restricted hypotheses must be arbitrary: the choice of restricted hypotheses cannot be justified by reference to observable implications. Thus, one must choose one set of restricted hypotheses over another for reasons that are unrelated to observable implications.

3 Implications of near observational equivalence for Bayesians

In the face of the sort of problems discussed here there are two natural ways to go for a solution. The first, restricting the hypotheses is probably not fruitful. The second is a Bayesian approach, using a prior that downweights problematic areas of the parameter space. Although Sims's [1971] work treated the Bayesian case, recent examination of Bayesian approaches by Sims and others (cited in the introduction) may have left impression that Bayesian work sidesteps the near observational equivalence problems faced by the classical statistician. For example, Sims and Uhlig [1991, p.1592] argue that "flat-prior Bayesian analysis leads to the usual t -tests for generating of posterior probabilities even in dynamic nonstationary models." Thus, they conclude [p.1599],

[T]he complicated apparatus of classical unit root asymptotics is of little practical value. Even econometricians who do not accept this conclusion, however, should agree that *the likelihood function's shape is valuable information*. (emphasis added)

¹² Sims [1972] makes similar arguments.

While statements such as these have rightly stimulated a great deal of discussion, this section points up a way in which both the claim and the responses are potentially misleading. In particular, the results below reconcile the Sims-Uhlig claim with the seemingly contradictory claim that the likelihood function's shape is not publicly informative about unit roots. The argument below is largely unrelated to earlier criticisms by Phillips [1991] and others. Much of that debate has centered on which prior one should use; thus, the debate is about how to interpret the information in the likelihood. The results below apply to *any* prior and suggest a particular sense in which the likelihood is uninformative.

3.1 Observational equivalence in Bayesian analysis

The issues raised in this section are best explained in terms of observational equivalence and several related Bayesian notions. Bayesian statisticians are not plagued by observational equivalence problems the same way that classical statisticians are. Consider a case in which there are some identified parameters and some unidentified ones. So long as the prior involves a correlation between the unidentified parameter and identified ones, then information that data provide about the identified parameter will, from the perspective of the person with such a prior, be informative about the unidentified parameter.

Observational equivalence does present Bayesian statisticians with a problem, however. In most simple cases with fully identified models, sufficient data will lead to consensus among a group of observers with non-dogmatic priors. In large enough samples, the data do all the talking, the priors become unimportant, and posteriors converge. This cannot be the case for unidentified parameters: the data do not speak at all on such subjects. As Leamer [1978] put it, data are not *publicly informative* about unidentified parameters: differences of opinion reflected in priors remain unaltered by data. One formalization of this statement is illustrated below.¹³

The tie between public informativeness and observational equivalence is made

¹³ Kadane [1974] and Drèze [1974] discuss in det the sense in which the influence of the prior remains.

clearer using Zellner's [1971] notion of observationally equivalent priors. Two priors are observationally equivalent if the prior probability of any observable events is the same under the two priors. If there are observationally equivalent priors that differ in their assessment of some parameter's value, then data will not be publicly informative about that parameter [Leamer, 1978].

A simple illustration may be helpful. The model is represented by the family of densities $p^X(X|\beta)$, and the priors are densities $p(\beta)$. Given r 's prior, p_r , and the model, the marginal prior density for any outcome is,

$$p_r^X(X) = \int p^X(X|\beta)p_r(\beta)d\beta$$

Two priors, q and r , are observationally equivalent if $p_q^X(X) = p_r^X(X)$ for all X . The posterior density for any β , conditional on observing a particular observation X , is given by

$$p_r(\beta|X) = \frac{p(X|\beta)}{p_r(X)}p_r(\beta)$$

If q and r agree on the model and have observationally equivalent priors, the first term in this expression will be the same; thus,

$$\frac{p_r(\beta|X)}{p_q(\beta|X)} = \frac{p_r(\beta)}{p_q(\beta)}$$

Any difference of opinion over β in the prior remains unaltered, in this proportional sense, in the posterior.

3.2 Definition of nearly observationally equivalent priors

As with the classical concept of observational equivalence, the related Bayesian notions can be loosened slightly to shed light on the TS/DS debate. In particular, this section shows that data are *nearly* not publicly informative about $A(1)$.

Return to notation for the maintained model used earlier. Assume ξ_r is r 's prior

probability measure on Θ .¹⁴ The marginal prior for any set of outcomes Γ is,

$$\xi_r^Y(\Gamma) = \int \mu_\theta(\Gamma) \xi_r(d\theta) \quad (3)$$

Definition 2 *The sequence of prior probability measures, $\{\xi_k\}$, has elements that are nearly observationally equivalent to ξ_r if for any fixed, finite, T , $\xi_k^Y \Rightarrow_k \xi_r^Y$, where \Rightarrow means convergence in distribution of the underlying random variables.*

Thus, two priors are nearly equivalent if the prior probability of observing events is arbitrarily close under the two priors. The definition of public informativeness can now be loosened to,

Definition 3 *Take the prior probability measure ξ_r and a sequence with nearly observationally equivalent priors, $\{\xi_k\}$. Data are nearly not publicly informative about θ if for all k , $\xi_k(\Theta_0) - \xi_r(\Theta_0) > \delta > 0$ for some measurable subset Θ_0 of Θ .*

In short, data are nearly not informative about a parameter if there are priors that agree to an arbitrary degree on the prior probability of any observation, but disagree absolutely on the prior probability of some parameters.

3.3 Data are nearly not publicly informative about $A(1)$

Given the definitions above, it is now relatively straightforward to show three statements about Bayesian analysis of $A(1)$:

Proposition 4 *Take any prior probability measure over the parameters of the maintained model. i) Data are nearly not publicly informative about $A(1)$. ii) There are nearly observationally equivalent priors that differ arbitrarily regarding the distribution of $A(1)$. iii) Given any observation, Y^T , there are observationally equivalent priors that give rise to posteriors differing arbitrarily regarding the distribution of $A(1)$.*

Proof: See the Appendix.

The intuition for the proof can be given for the case in which the prior is represented as a density. Suppose one prior density takes the value g at θ . Construct

¹⁴ Various measure theoretic details about this statement are dealt with more carefully in the Appendix. There are well-known problems associated with putting probability measures on infinite dimensional parameter spaces and with Bayesian inference in such cases [e.g., Sims, 1971, and Diaconis and Freedman, 1986]. The problem discussed in this paper is closely related.

the observationally equivalent prior by putting prior density g on some θ_k that is nearly observationally equivalent to θ . This is done for each θ . Whatever one prior says about θ the other says about a nearly observationally equivalent θ_k . Since θ and θ_k have similar implications, the two priors will agree on observable outcomes. The θ_k s, however, can have quite different long-run effects of shocks than the θ s.

Prop. 4 sheds some light on the discussion of the Sims-Uhlig quote above. They claim that the likelihood provides useful information about the parameters of the model. In contrast, prop. 4 states that the data contain no public information about unit roots. These two views are easily reconciled.

The sort of *flat prior* analysis advocated by Sims and Uhlig must restrict the prior to put mass only on a small set of MA representations. No proper posterior—that is, no probability measure—can put mass more than a meager set of l^1 [Parthasarathy 1967]. Thus, the implicit *flat prior* in the analysis is similarly restricted. Prop. 4 establishes that there are other *flat priors* over a different portion of the parameter space that will treat observable outcomes in the same manner as the first *flat prior*, but that give rise to an arbitrarily different posterior for unit root parameters such as $A(1)$. Thus, the choice of a support for the *flat prior*—the choice of a small set of parameters over which to investigate the likelihood—determines the answer regarding unit root issues.

The Sims-Uhlig critique of classical procedures and the Phillips response carry their full weight once both sides have agreed to restrict all analysis to one small portion of the DS/TS parameter space—as opposed to some other similarly small portion.¹⁵ However, if that restriction is essentially arbitrary and determines the answer to unit root questions, one must be quite careful in interpreting the debate. In particular, any statement that the likelihood is *informative* about unit roots or even a statement suggesting that one form of inference regarding unit roots is *better* than another is suspect. The answer obtained under arbitrary restrictions is

¹⁵ Of course, much of the debate has focussed strictly on autoregressive structures of order one.

similarly arbitrary, no matter what the approach.¹⁶

Overall, prop. 4 closely circumscribes the cases under which reported posterior probabilities regarding the presence of unit roots are of interest. Even if one agrees that the prior reported in some empirical work correctly captures the prior probability of observable outcomes, the posterior over unit root issues may be irrelevant: there will be another prior treating outcomes in the same manner, but for which the posterior is arbitrarily different. The prior over outcomes has almost no implications for the prior (and, hence, the posterior) over unit root issues. Reported results are of interest only if one agrees with aspects of the prior over parameters that are independent of empirical implications. In practice, I believe we have little or no basis for selection of such a prior.

4 Extension to the multivariate case

The results of this paper all generalize easily to the multivariate case. The maintained model generalizes by simply reinterpreting Y_t as an $N \times 1$ vector. In the Wold representation, $A(L)$ is now matrix of lag polynomials, $\sum_{i=0}^{\infty} a_i L^i$, where a_i is $(N \times N)$. The long-run effect matrix $A(1)$ is now $(N \times N)$.

Of course, rather than simply asking whether the process is DS or TS we must consider cointegration: are any linear combinations of the variables stationary? By direct analogy with the univariate case, one can easily show that if $w'A(1) = 0$, then this linear combination of the variables is TS. In this case, of course, the variables are called cointegrated with cointegrating vector w .¹⁷ Of course, elementary linear algebra tells us that if $w'A(1) = 0$ then $A(1)$ is singular; there can be at

¹⁶ Some care should be taken here. The maintained model in this paper does not allow for explosive roots, which have played a central role in the Phillips critique. This restriction is probably a strength of the results in this paper, since they show that inference regarding stationarity is fragile, even if one rules out explosive roots. Further, when explosive roots are allowed, the argument still applies to the portion of the parameter space consistent with the DS model, and, hence, shows that any posterior mass attributed to TS structures could just as well be transferred to the nonstationary portion of the parameter space. More direct treatment of the explosive case is beyond the scope of this paper.

¹⁷ Details about deterministic elements are ignored here, as in the univariate results.

most N such linearly independent vectors w ; if there are j such vectors, then the rank of $A(1)$ is $n - j$; j is the cointegrating rank of the system. Thus, the rank- j cointegration hypothesis can be stated

$$H_{CI(j)} : \text{rank}(A(1)) = n - j \quad \text{for some } 0 \leq j \leq n$$

With these changes understood, all the results of the paper generalize directly. Prop. 1 is proven for the general case in the Appendix. The remaining props. follow as well: any problem distinguishing TS from DS structures in the univariate case carries over to distinguishing $H_{CI(j)}$ structures from $H_{CI(k)}$ structures in the multivariate case.

5 Conclusions

Difference stationary and trend stationary models are nearly observationally equivalent. Just as with strict observational equivalence, any classical inference regarding unit roots must be an artifact of implicit or explicit restriction. Similarly, any Bayesian conclusion is fragile to changes in the prior that have (almost) no implications for observable outcomes. The restrictions needed to resolve these problems for the classical or Bayesian statistician cannot be justified by reference to empirical plausibility. They must be grounded in some sort of *a priori* reasoning.

Despite near observational equivalence of DS and TS models, there is, in my experience, a strong tendency among econometricians to believe that certain empirical behavior *typifies* DS processes and that other behavior *typifies* TS processes. In this perspective, the DS processes that behave like *typical* TS processes are peculiar in some sense, and can safely be ignored or ruled out. The assignment of *typical* behavior in this case is hazy at best and arbitrary and misleading at worst. Absent some clear elucidation of the conditions under which such thinking is justified, one is probably on safer ground with the view that there is no practical distinction between the behavior of DS and TS structures.

For economists, acceptance of this view implies that one cannot distinguish theories by deriving and testing implications regarding the long-run effect of shocks. For practicing econometricians, this means that no test statistic should be thought to behave differently depending only on whether the underlying process is DS or TS. While the behavior of standard test statistics may differ with *certain aspects* of the dynamics of the underlying model, those certain aspects are distinct from unit root questions. As Blough [1991] has emphasized, profitable work in this regard may come from looking beyond the TS/DS distinction in attempting to resolve the inference problems that have been highlighted in the unit root literature.

Appendix

Proof of prop. 1. The proof is done for the multivariate case, $Y^T(T \times N)$. The distribution of Y^T is described by $\theta = \{A, \psi\}$, where A is $(N \times N)$. Take the sequence of random variables $\{Y_k^T\}$ described by $\theta_k = \{A_k, \psi\}$, where $\|A_{ij} - A_{kij}\|_2 \rightarrow_k 0$ for each i, j , where A_{ij} is the sequence of coefficients of the scalar polynomial $[A(L)]_{i,j}$. Such a sequence can be constructed satisfying arbitrary rank and eigenvalue restrictions in the discussion following prop. 1.

A sufficient condition for convergence in distribution of Y_k to Y is that every finite linear combination of the Y_k converges in mean square to the similar linear combination of Y :

$$E\left(\sum_{t=1}^T \sum_{n=1}^N w_{tn}(Y_{tn} - Y_{tnk})\right)^2 \rightarrow_k 0 \quad (4)$$

The double sum has mean zero, so the term is a variance. The double sum can be re-stated as a weighted sum of terms $Z_{njtk} = (A_{nj}(L) - A_{njk}(L))\nu_{nt}$ where the number of terms, M , is fixed by T and N and the weights are fixed by T , N , and the w_{tn} . The variance of such a term is bounded by M^2 times the maximum weight squared times the maximum variance of any Z term. Since M and the maximum weight are fixed independent of k , it is sufficient to show that the variance of all the Z terms goes to zero with k . Of course, $\text{var}(Z_{njtk}) = \|A_{nj} - A_{njk}\|_2^2 \text{var}(\nu_n)$, which goes to zero by assumption. Q.E.D.

Proof of prop. 3. Initially consider part *ii*: Take either hypothesis as the null, and define the rejection region of the test procedure in a sample of size T to be Γ_T . By assumption there is some structure r , consistent with the alternative for which, $\lim_{T \rightarrow \infty} \mu_{\theta_r}^T(\Gamma_T) = 1$. By near observational equivalence, for any fixed sample size T , there exists a sequence of structures parameterized by $\{\theta_k^T\}$, consistent with the null, such that

$$\lim_{k \rightarrow \infty} \mu_{\theta_k^T}^T(\Gamma_T) - \mu_{\theta_r}^T(\Gamma_T) = 0 \quad (5)$$

Since this relation holds for all T , we can choose one element from each sequence,

$\{\theta_k^T\}$, $T = 1, 2, \dots$, to form a new sequence $\{\theta_T\}$ satisfying

$$\lim_{T \rightarrow \infty} \mu_{\theta_T}^T(\Gamma_T) - \mu^T(\Gamma_T | \theta_r) = 0$$

and, hence, satisfying $\mu(\Gamma_T | \theta_T) \rightarrow 1$.

Part *i* can be proven by fixing T , and realizing that for any θ_r consistent with the alternative there is a sequence $\{\theta_k\}$ satisfying (5). Q.E.D.

Proof of prop. 4. Parts *i* and *ii*: First, lay out some technical detail to make sure various objects are well-defined. A prior is a probability measure on some parameter space and a mapping from that space to probability distributions for the data. Assume that the parameter space, Θ^* , is a Borel subset of a complete separable metric space. To each $\theta \in \Theta^*$ there is associated a probability measure on the Borel sets of R^T consistent with the maintained model. Assume that the map, $H : \theta \rightarrow \mu_\theta$ is one-to-one and Borel. Under these assumptions, the standard joint prior over Y^T and θ can be derived, and if ξ is a probability measure on Θ^* , the marginal prior for observable events is given by (3). For existence of the prior for $A(1)$, assume that for Borel sets, B , on the real line, the sets $S(B) = \{\theta : s\theta \in B\}$ are ξ -measurable, where $s\theta$ is the long-run effect of shocks for the parameter θ .

Given any prior over Θ^* defined by ξ and the map H , each member of the sequence of observationally equivalent priors is formed by choosing a one-to-one mapping, h_k from Θ^* onto some subset Θ_k^* of l^1 . The prior probability measure, then, is defined by $\xi_k(\Theta_0) = \xi(h_k^{-1}(\Theta_0))$ for measurable subsets Θ_0 of Θ_k^* (the Borel sets of Θ_k^* are generated by h_k from those of Θ^*). Now define $h_k : \theta$ and $h_k(\theta)$ have the same ψ . They differ in their MA representations in that A_k has z/k added to its first k elements for an arbitrary scalar z . Under this formulation, the new priors can be represented as the prior ξ on Θ^* and the mapping $H_k : \theta \rightarrow \mu_{h_k(\theta)}$.

By prop. 1, for each θ , $\mu_{h_k(\theta)}(\Gamma) \rightarrow_k \mu_\theta(\Gamma)$ for μ_θ -continuity sets Γ . That is, for fixed Γ , $\mu_{h_k(\theta)}$ converges pointwise in θ to μ_θ . Since $\mu_\theta(\Gamma) \in [0, 1]$, standard general convergence theorems apply (e.g., Royden 1968, prop. 18, p.232), and

$$\xi_k^Y(\Gamma) = \int_{\Theta^*} \mu_{h_k(\theta)}(\Gamma) \xi(d\theta) \rightarrow_k \int_{\Theta^*} \mu_\theta(\Gamma) \xi(d\theta) = \xi^Y(\Gamma)$$

for continuity sets Γ . This implies $\xi_k^Y \Rightarrow \xi^Y$.

Now show that the ξ_k can treat $A(1)$ arbitrarily different from ξ . By arbitrarily different we mean that if ξ puts mass $1 - \varepsilon$ on $A(1)$ being in some set, there are nearly observationally equivalent priors that put mass less-than-or-equal-to ε on $A(1)$ being in that set. Take an arbitrary $0 < \varepsilon < 1$, and pick any interval, $[b, c]$, such that $\xi(S([b, c])) = 1 - \varepsilon$. By construction, each member of the nearly observationally equivalent sequence assigns mass $1 - \varepsilon$ to the interval $[b + z, c + z]$. Thus, setting $z = c - b$, the equivalent priors put mass $1 - \varepsilon$ to an interval outside $[b, c]$, and must assign less than or equal to ε to $[b, c]$.

Part *iii* follows directly. Q.E.D.

References

- Berk, Kenneth. "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 1974, 2, pp.489–502.
- Blough, S. R. "Near Observational Equivalence of Unit Root and Stationary Processes," manuscript, Johns Hopkins University, December 1989.
- _____. "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples," manuscript, John Hopkins University, 1991.
- Campbell, John Y., and Pierre Perron. "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," Technical Working Paper 100, Cambridge, MA: National Bureau of Economic Research, 1991.
- Christiano, L. J., and M. Eichenbaum. "Unit Roots in Real GNP: Do We Know, and Do We Care?," *Carnegie-Rochester Conference Series on Public Policy*, 1990, 32, pp. 7–62.
- Cochrane, J. H. "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control*, 1991, 15, pp. 275–284.
- DeJong, David, and Charles Whiteman, "Unit Roots in U.S. Macroeconomic Time Series," in *New Directions in Time Series Analysis*, Brillinger et.al., eds., IMA volumes in mathematics and its applications, vol. 46, Springer Verlag, 1993, pp. 43–69.
- Diaconis, Persi, and David Freedman. "On the Consistency of Bayes Estimates," *Annals of Statistics*, 1986, pp. 1–26.
- Dickey, D. A., and W. A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74, 1979, pp. 427–431.
- Drèze, Jacques. "Bayesian theory of identification in simultaneous equations models," in *Studies in Bayesian Econometrics and Statistics*, Fienberg and Zellner eds., New York: North Holland, 1974, pp.159–174.
- Hall, A. "Testing for a Unit Root in the Presence of Moving Average Errors," *Biometrika*, 1988.
- Kadane, Joseph. "The roole of identification in Bayesian theory," in *Studies in Bayesian Econometrics and Statistics*, Fienberg and Zellner eds, 1974, pp.175–191.
- Leamer, E. *Specification searches: ad hoc inference with non-experimental data*, New York: John Wiley, 1978.
- Oxtoby, John. *Measure and Category*, New York: Springer-Verlag, 1980.

- Pantula, Sastry G. "Asymptotic Distribution of the Unit Root Tests When the Process is Nearly Stationary," North Carolina State University, 1988.
- Pathasarathy, K.R. *Probability Measures on Metric Spaces*, New York: Academic Press, 1967.
- Phillips, P.C.B. "To criticize the critics: an objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 1991, 3, pp.333-364.
- Royden, H.L. *Real Analysis*, New York: MacMillan, 1968.
- Schwert, G. W. "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 1989, 7, pp. 147-160.
- Sims, Christopher. "Distributed lag estimation when the parameter space is explicitly infinite-dimensional," *Annals of Mathematical Statistics*, 1971, 42, pp.1622-1636.
- _____. "The role of approximate prior restrictions in distributed lag estimation," *Journal of the American Statistical Association*, 1972, 67, pp. 169-175.
- _____. "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control*, 1988, 12, pp.463-474.
- _____. "Comment on 'To criticize the critics', by Peter C. B. Phillips," *Journal of Applied Econometrics*, 1991, 3, pp. 423-434.
- _____, and Harald Uhlig, "Understanding unit rooters: a helicopter tour," *Econometrica*, 1991, 59, pp.1591-1599.
- Zellner, Arnold. *An introduction to Bayesian inference in econometrics*, New York: Wiley, 1971.

International Finance Discussion Papers

IFDP
NUMBER

TITLES

AUTHOR(s)

1993

- | | | |
|-----|------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| 447 | Near observational equivalence and unit root processes: formal concepts and implications | Jon Faust |
| 446 | Market Share and Exchange Rate Pass-Through in World Automobile Trade | Robert C. Feenstra
Joseph E. Gagnon
Michael M. Knetter |
| 445 | Industry Restructuring and Export Performance: Evidence on the Transition in Hungary | Valerie J. Chang
Catherine L. Mann |
| 444 | Exchange Rates and Foreign Direct Investment: A Note | Guy V.G. Stevens |
| 443 | Global versus Country-Specific Productivity Shocks and the Current Account | Reuven Glick
Kenneth Rogoff |
| 442 | The GATT's Contribution to Economic Recovery in Post-War Western Europe | Douglas A. Irwin |
| 441 | A Utility Based Comparison of Some Models of Exchange Rate Volatility | Kenneth D. West
Hali J. Edison
Dongchul Cho |
| 440 | Cointegration Tests in the Presence of Structural Breaks | Julia Campos
Neil R. Ericsson
David F. Hendry |

1992

- | | | |
|-----|--------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|
| 439 | Life Expectancy of International Cartels: An Empirical Analysis | Jaime Marquez |
| 438 | Daily Bundesbank and Federal Reserve Intervention and the Conditional Variance Tale in DM/\$-Returns | Geert J. Almekinders
Sylvester C.W. Eijffinger |
| 437 | War and Peace: Recovering the Market's Probability Distribution of Crude Oil Futures Prices During the Gulf Crisis | William R. Melick
Charles P. Thomas |
| 436 | Growth, Political Instability, and the Defense Burden | Stephen Brock Blomberg |
| 435 | Foreign Exchange Policy, Monetary Policy, and Capital Market Liberalization in Korea | Deborah J. Lindner |

Please address requests for copies to International Finance Discussion Papers, Division of International Finance, Stop 24, Board of Governors of the Federal Reserve System, Washington, D.C. 20551.

International Finance Discussion Papers

<u>IFDP NUMBER</u>	<u>TITLES</u>	<u>AUTHOR(s)</u>
<u>1992</u>		
434	The Political Economy of the Won: U.S.-Korean Bilateral Negotiations on Exchange Rates	Deborah J. Lindner
433	Import Demand and Supply with Relatively Few Theoretical or Empirical Puzzles	Andrew M. Warner
432	The Liquidity Premium in Average Interest Rates	Wilbur John Coleman II Christian Gilles Pamela Labadie
431	The Power of Cointegration Tests	Jeroen J.M. Kremers Neil R. Ericsson Juan J. Dolado
430	The Adequacy of the Data on U.S. International Financial Transactions: A Federal Reserve Perspective	Lois E. Stekler Edwin M. Truman
429	Whom can we trust to run the Fed? Theoretical support for the founders views	Jon Faust
428	Stochastic Behavior of the World Economy under Alternative Policy Regimes	Joseph E. Gagnon Ralph W. Tryon
427	Real Exchange Rates: Measurement and Implications for Predicting U.S. External Imbalances	Jaime Marquez
426	Central Banks' Use in East Asia of Money Market Instruments in the Conduct of Monetary Policy	Robert F. Emery
425	Purchasing Power Parity and Uncovered Interest Rate Parity: The United States 1974 - 1990	Hali J. Edison William R. Melick
424	Fiscal Implications of the Transition from Planned to Market Economy	R. Sean Craig Catherine L. Mann
423	Does World Investment Demand Determine U.S. Exports?	Andrew M. Warner
422	The Autonomy of Trade Elasticities: Choice and Consequences	Jaime Marquez
421	German Unification and the European Monetary System: A Quantitative Analysis	Gwyn Adams Lewis Alexander Joseph Gagnon
420	Taxation and Inflation: A New Explanation for Current Account Balances	Tamim Bayoumi Joseph Gagnon