The Power of the "Objective" Bayesian Unit-Root Test

Francis W. Ahking[*]

Department of Economics, U-63, University of Connecticut, Storrs, CT 06269-1063, USA

Abstract: Some researchers, for example, Koop [1], and Sims [2], have advocated for Bayesian alternatives to unit-root testing over the classical approach using the augmented Dickey-Fuller test (ADF). This paper studies the power of what Koop [1] has called the "Objective" Bayesian approach to unit-root testing. Koop's "objective" Bayesian test is interesting in light of the call by Phillips [3, 4] for more objective Bayesian analysis of time series. We apply the "objective" Bayesian unit-root test to a study of long-run purchasing power parity (PPP) in the post-Bretton Woods era and also Monte Carlo simulations. Overall, contrary to the favorable simulation results obtained by Koop [1], our results suggest that the "objective" Bayesian test is biased in favor of trend-stationarity. We conclude that, at least for the "objective" Bayesian test, it is not better than the classical ADF approach in unit-root tests, and because of its bias, the "objective" priors suggested by Koop [1] is not appropriate, and we do not recommend its use.

JEL Classification: C11, C22, F31.

1. INTRODUCTION

The question of how best to characterize the growth component of macroeconomic time series has been hotly debated since the publication of the study by Nelson and Plosser [5]. The most widely held view before 1982 was that most macroeconomic time series could be characterized as trend-stationary (TS) series. Nelson and Plosser's study called into question that widely held view by demonstrating that thirteen out of fourteen USA macroeconomic time series in their study contained a unit root, and thus were difference-stationary (DS) series. Although the debate is far from settled,[1] a related debate is concerned with the statistical techniques to use in testing between TS and DS series. Several economists, e.g., DeJong and Whiteman [7], Koop [1], and in particular, Sims [2], and Sims and Uhlig [8], have advocated forcefully for Bayesian alternatives over the more traditional classical approach such as the augmented Dickey-Fuller (ADF) tests [9] in unit-root testing. These economists cited several advantages of the Bayesian approach over the classical approach. For example, it is well known that ADF tests have low power against plausible alternatives, especially against trend-stationary alternative. The Bayesian approach, on the other hand, would reveal that both the unit root and the trend-stationary hypotheses would receive similar posterior probabilities. Thus, the Bayesian approach provides a more reasonable summary of sample information than the classical approach. Another problem with the classical unit-root tests is the discontinuity of the classical asymptotic theory when there is a unit root (see Sims [2]

especially for a good discussion of this point). The Bayesian approach, on the other hand, since it is based on the likelihood function, does not have the same discontinuity problem. Finally, Koop [10] has also pointed out that, in the classical approach, the small sample critical values, which are used frequently, could very often differ substantially from the asymptotic critical values. The Bayesian approach, on the other hand, since it is conditional on the observed sample, provides exact small sample results. Indeed, an entire issue of the Journal of Applied Econometrics [11] was devoted to a discussion of these and other issues.

Despite the apparent advantages of the Bayesian approach over the classical approach in unit-root testing, only a relatively small number of studies have appeared using the Bayesian approach. The reasons may be that the Bayesian approach requires a likelihood function and the use of prior information. The perception is that other than for the simplest cases, the Bayesian approach is computationally burdensome, and the use of prior information is by far the most controversial. Phillips [3, 4] identified the need for priors as the biggest obstacle to Bayesian analysis and argued for more objective Bayesian analysis in time series.

In this paper, we study what Koop [1] has called the "objective" Bayesian test. Koop's [1] Bayesian unit-root test is based on the work of Zellner and Siow [12]. There are several features of this approach that make it attractive for economists looking for an alternative to the classical unit-root tests. First, it is computationally simple. Second, it requires informative priors, thus avoiding the problems associated with using non-informative priors, which are frequently improper. Third, it does not require significant subjective prior information, only that all competing hypotheses have equal prior probability, leading Koop [1] to call this an "objective" Bayesian test.

*Address correspondence to this author at the Department of Economics, U-63, University of Connecticut, Storrs, CT 06269-1063, USA; Tel: (860) 486-3026; Fax: (860) 486-4463; E-mail: Francis.Ahking@uconn.edu

---

[1] For a recent contribution to the debate, see Murray and Nelson [6].

Koop's [1] Bayesian test appears to address one of the biggest criticisms of Bayesian econometrics and is very much in the spirit of Phillips's [3, 4] call for more objective analysis in Bayesian time series. However, the reliability of Koop's [1] Bayesian unit-root test, and in particular, whether or not it provides a better alternative to classical unit root tests, and how appropriate are his objective priors are questions that have not been examined extensively. Koop [1] himself has provided some limited simulation results on the size and power of the "objective" Bayesian unit-root test, and has obtained mostly favorable results. The one major drawback however is that Koop used a data generating model (DGM) which is identical to the one used by DeJong, Nankervis, Savin, and Whiteman [13], and test statistics of this DGM are dependent on the initial starting value.[2] There is no optimal way to determine the appropriate starting value, however. In addition, we are not aware of any other published studies that have examined the power of the "objective" Bayesian unit-root test. Thus, the reliability of the "objective" Bayesian unit-root test is very much an opened question. In this paper, we seek to study the two related questions of reliability and the appropriateness of the objective priors by providing a Monte Carlo study of the power of Koop's [1] "objective" Bayesian unit-root test. We will use a less restrictive DGM and will also conduct our simulations using different sample sizes and parameter values. We start our empirical analysis by applying the "objective" unit-root test to a study of long-run purchasing power parity (PPP) to see how this test would perform in actual empirical application. This is also motivated by the fact that frequently, the usefulness of an empirical test is determined, to a large extent, by how well it performs in actual application. Next, we perform a number of simulations to assess the power of the test and thus provide information on both the usefulness of the test and the appropriateness of the "objective" priors.

As we will argue below (see footnote 4), however, eventhough we assign equal prior probability to all the competing hypotheses, the unit-root hypothesis actually receives the highest prior probability. We provide a brief discussion of Koop's [1] "objective" Bayesian unit-root test in the next section. We present our empirical results on long-run PPP in section 3. We study the properties of the "objective" Bayesian unit-root test using Monte Carlo simulations in section 4. Finally, our summary and conclusions are in section 5.

2. KOOP'S "OBJECTIVE" BAYESIAN APPROACH

The modeling objective of the Bayesian approach is not to reject or fail to reject a hypothesis based on a pre-determined level of significance, but to determine how probable a hypothesis is relative to other competing hypotheses. There are several ways of comparing hypotheses

using Bayesian methods. The most common method is to calculate posterior odds ratios for various competing hypotheses based on prior and sample information. This gives the researcher the odds in favor of one hypothesis relative to other competing hypotheses. The objective in Koop's "objective" Bayesian approach is to find the linear model that would best describe the time series, $\underline{q} = (q_1, q_2, ..., q_T)'$. We consider three hypotheses:

$$H_1: \quad q_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i q_{t-i} + \alpha_{k+1} t + \xi_{1t}, \quad (1)$$

$$H_2: \quad q_t = \alpha_0 + \sum_{i=1}^{k} \alpha_i q_{t-i} + \xi_{2t}, \quad (\alpha_{k+1} = 0) \quad (2)$$

$$H_3: \quad \begin{aligned} \Delta q_t = \alpha_0 - (\sum_{i=2}^{k} \alpha_i)\Delta q_{t-1} - \\ (\sum_{i=3}^{k} \alpha_i)\Delta q_{t-2} - \cdots - \alpha_k \Delta q_{t-(k-1)} + \xi_{3t} \\ (\sum_{i=1}^{k} \alpha_i = 1, \alpha_{k+1} = 0) \end{aligned}, \quad (3)$$

where $t=$ a linear deterministic time trend; and $\xi_{jt}, j = 1, 2, 3$ is a serially uncorrelated error process with zero-mean and constant variance.

Hypothesis 1 (H_1) is the null model. It hypothesizes a trend-stationary auto-regressive process of order k, i.e., a trend-stationary AR(k) process. H_2 hypothesizes a stationary AR(k) process, while H_3 hypothesizes an AR(k) process with a unit root. Note that both H_2 and H_3 are special cases of H_1 with linear restrictions (given in parentheses next to the respective equations), imposed on the null model. The trend-stationary hypothesis is included because it is the leading alternative to unit-root non-stationarity in macroeconomics time series. The stationary alternative is also included to see how well the "objective" Bayesian test can distinguish between non-stationary series and stationary series with a high degree of persistence, as is frequently encountered in macroeconomic time series.

We compare the three hypotheses, based on both prior and sample information, by calculating the posterior odds ratios:

$$K_{1j} = \frac{P(H_1)P(H_1 | \underline{q})}{P(H_j)P(H_j | \underline{q})}, \quad j = 2, 3, \quad (4)$$

where $P(H_1)/P(H_j)$ is called the prior odds ratio, and $P(H_i | \underline{q}) = \int P(\theta_i | H_i)L(\theta_i | \underline{q}, H_i)d\theta_i, i = 1, 2, 3$, is the posterior probability that $H_i, i = 1, 2, 3$, were true given the sample data $\underline{q}$, $P(\theta_i | H_i)$ and $L(\theta_i | \underline{q}, H_i)$ are the prior density for θ_i, and likelihood function, respectively, under

---

[2] Specifically, the test statistics is dependent on the "standardized initial displacement", which, according to DeJong, Nankervis, Savin, and Whiteman [13], "measures the distance (in units of the innovation standard deviations) between the initial value y_0 and the trend line", where y_0 is the initial value of the generated time series y_t.

each hypothesis, and $\underline{\theta}$ is a vector of parameters. Thus, the posterior odds ratio gives the ratio of the probabilities of the two hypotheses being true given the sample data.[3] On the assumption that all three hypotheses have equal prior probability,[4] i.e., $P(H_1) = P(H_2) = P(H_3)$, equation (4) reduces to

$$K_{1j} = \frac{P(H_1 \mid \underline{q})}{P(H_j \mid \underline{q})}, \; j = 2, 3 \, . \tag{5}$$

Following Koop [1], we calculate the posterior odds ratio for testing a set of exact linear restrictions with a formula suggested by Zellner and Siow [12]. The Zellner-Siow posterior odds ratio is calculated approximately as:

$$K_{1j} \cong \frac{(\pi^{0.5} / \Gamma[(r+1)/2])(v/2)^{r/2}}{(1 + rF/v)^{(v-1)/2}}, \; j = 2, 3 \, , \tag{6}$$

where $\Gamma[\,\cdot\,]$ = the Gamma function, $v = T - n$, T = the total number of observations, n = the number of regressors in the null model, r = the number of linear restrictions tested, and F = the usual F-statistics for testing the set of linear restrictions. After we obtained the posterior odds ratios, we can then calculate the posterior probability for each of the three hypotheses. We apply the "objective" Bayesian test to a study of long-run PPP in the next section.

3. TESTING FOR LONG-RUN PPP

The theory of Purchasing Power Parity (PPP) occupies a central place in international economics. It is a key building block of the monetary models of exchange rate determination. In the flexible-price monetary model of exchange rate (e.g., Frenkel [14]), PPP is assumed to hold continuously. In the sticky-price monetary model (e.g., Dornbusch [15]), because of sticky prices in the short run, PPP does not hold, but is a maintained assumption for the long run. Recent empirical tests of PPP have mainly focused on the long run given that there are frequent large and persistent deviations from PPP. The question of interest is whether deviations from PPP are transitory or permanent. Thus, the empirical tests generally take the form of testing for stationarity of the real exchange rate. If deviations from PPP are transitory, we would expect the time series of the real exchange rate to be stationary. In this case, PPP holds for the long run. On the other hand, the finding of a unit-root non-stationary time series of the real exchange rate would

imply that deviations from PPP are permanent, and PPP does not hold.

There is a vast empirical literature on long-run PPP using a variety of empirical techniques and approaches.[5] Examples from recent studies include Diebold, Husted, and Rush [18] using the autoregressive fractionally integrated moving-average approach, Engel and Kim [19] using a long span of data that mixed both fixed and flexible exchange rate data, O'Connell [20] using panel data, Murray and Papell [21] using a structural-break adjusted approach, and Ahking [17] using a non-parametric approach. Unfortunately, the results have been largely inconclusive and also not very robust, especially using data from the post-Bretton Woods era (see the surveys by Rogoff [22], and Sarno and Taylor [16]). Only a few empirical studies have used a Bayesian approach, however (see, for example, Schotman and Van Dijk [23], Whitt [24], and Ahking [25]). Given the alleged advantages of using the "objective" Bayesian unit-root test discussed above, we present in this section a study of long-run PPP using the "objective" Bayesian unit-root approach.

We define the real exchange rate in natural logarithm form as:

$$q_t = e_t + P_t^* - P_t \, ,$$

where q_t is the natural logarithm of the real exchange rate, e_t is the natural logarithm of the nominal exchange rate, defined as the domestic currency price of one unit of foreign currency, and P_t (P_t^*) is the natural logarithm of an index of the domestic (foreign) price level. A test for long-run PPP is a test of whether or not q_t is a stationary time series. The source of our data is SourceOECD, and consist of the OECD G-7 countries - the U.S., the U.K., Canada, Germany, Italy, France, and Japan. Our data consist of monthly observations from April 1973 to February 1999 for the G-7 countries, and are not seasonally adjusted. The sample ends when several of the G-7 countries became part of the Euro zone. Nominal exchange rates are bilateral and are monthly averages. In all cases, we use the consumer price index as our measure of the average price level. Thus, we have a total of twenty-one real exchange rates in our sample.[6]

We start our empirical tests by first presenting our test for unit-root using the ADF test. The ADF unit-root test results will provide a comparison to our Bayesian approach to unit-root testing. The ADF regression actually estimated is:

$$\Delta q_t = \beta_0 + \phi q_{t-1} + \sum_{i=1}^{l} \beta_i \Delta q_{t-i} + \beta_{l+1}t + v_t \, , \tag{7}$$

---

[3] Koop [1] also considered what he called the g-prior approach. This approach, however, requires the subjective input by the researcher on the g values, and thus is less "objective" than the approach considered here.

[4] One could argue that even-though the three competing hypotheses have equal prior probability, the unit-root hypothesis is the most favored, and the trend-stationary hypothesis is the least favored. For example, the prior probability for the unit-root hypothesis, i.e., $\rho = 1$ is 33.33 percent, while the stationary interval $(0 < \rho < 1)$ also receives 33.33 percent prior probability, distributed uniformly in that interval, resulting in each point in that interval receiving extremely low prior probability. For the trend-stationary hypothesis, however, the 33.33 percent prior probability, in addition to being distributed on the stationary interval, must also be distributed on the infinite interval of values that the time-trend coefficient may take.

---

[5] Sarno and Taylor [16] provided a good survey on the recent literature on long-run PPP and a summary of the more recent empirical results on long-run PPP can be found in Ahking [17].

[6] The bilateral nominal exchange rates available have the U.S. dollar as the base currency, i.e., foreign currency per U.S. dollar. Non-U.S. dollar based exchange rates are computed as cross-rates. This assumes cross-rate equality except for transaction costs. This is probably a valid assumption for the G-7 countries. Alternatively, as long as the measurement error is a stationary process, our tests for unit-root will not be affected.

where t = a linear deterministic time trend, and v_t is a serially uncorrelated error process with zero mean and constant variance. We have included a linear deterministic time trend since several researchers have found that the stochastic processes of some of the real exchange rates in our study could not be adequately modeled without the inclusion of a linear deterministic time trend (see Cheung and Lai [26], and Koedijk, Schotman, and Van Dijk [27] for examples). Furthermore, since the time series of real exchange rates typically show evidence of a trend, the trend-stationary hypothesis offers an alternative to the unit-root hypothesis as the source of the non-stationarity in the data. The linear deterministic time trend in the real exchange rate is generally interpreted as representing systematic differences in productivity growth between tradable and non-tradable goods in the two countries (see Cheung and Lai [26] for a discussion). Thus, the presence of a linear deterministic time trend in the real exchange rate time series is generally not interpreted as a violation of long-run PPP.[7] The lag length for the lagged first-differences is determined by using a general-to-specific method recommended by Ng and Perron [30] and Perron [31]. We start by estimating Equation (7) with a pre-determined maximum lag length of 12 and sequentially drop the last included lag if it is not statistically significant at the 10% significance level. If, however, the lag length determined is the same as the maximum lag length, we start over with a maximum lag length of 14.[8]

Table **1** reports the empirical results using the ADF test where we also report results of estimating Equation (7) without a linear time trend. Column 2 reports the lag length selected for each real exchange rate, columns 3 and 4 report the t-statistic for the hypothesis $H_0 : \phi = 0$ without and with a trend, respectively. We use the 5% and the 10% critical values from Fuller [32] and the lag-adjusted critical values for exact sample size from Cheung and Lai [33]. Long run PPP would require the rejection of the null hypothesis. From Table **1**, column 3, we see that there are only three cases out of twenty one that the null hypothesis is rejected when Equation (7) is estimated without a trend, and five cases out of twenty one (column 4) when a trend is included. The results are as expected and consistent with earlier other studies using the ADF tests. The evidence against long-run PPP is rather strong without the inclusion of a time trend and slightly better when a time trend is included.

Table **2** presents the results of the "objective" Bayesian tests. In order to focus on the three main hypotheses, we use a non-Bayesian method to determine the lag length.[9] Since it

is not clear how best it is to determine the lag length for an equation such as equation (1), we keep it simple by starting with a lag length of one, and then two, and so on, and stopping when we obtain a white-noise error process.[10] We also experimented with using a uniform lag of four for all the real exchange rates, the results are consistent with those reported in Table **2**.

Table 1.   Univariate ADF Test Results

Real Exchange Rate	l	$t(\phi)$ without Trend	$t(\phi)$ with Trend
Canada/U.S.	12	-1.373	-2.025
U.K./U.S.	11	-2.594**	-2.725
Germany/U.S.	10	-2.331	-2.472
Italy/U.S.	10	-2.507	-2.545
Japan/U.S.	11	-1.932	-2.731
France/U.S.	3	-2.249	-2.244
U.K./Canada	11	-2.014	-2.719
Germany/Canada	10	-2.642**	-2.660
Italy/Canada	1[+]	-2.057	-3.298**
Japan/Canada	11	-1.779	-3.416**
France/Canada	5	-2.329	-2.663
Germany/U.K.	12	-2.196	-2.793
Italy/U.K.	11	-2.356	-2.351
Japan/U.K.	10	-2.035	-2.595
France/U.K.	1	-2.273	-2.630
Italy/Germany	2	-1.654	-1.917
Japan/Germany	8	-1.668	-3.598*
France/Germany	9	-3.066**	-4.499*
Japan/Italy	3	-2.433	-3.101
France/Italy	3	-1.966	-1.991
France/Japan	8	-1.983	-3.791*

Note: *,** Denote the rejection of the null hypothesis at the 5% and the 10% significance levels, respectively.
+ The lag length is 10 for the Canadian dollar/lira real exchange rate when estimated without a linear time trend.

What we find striking is that, of the three hypotheses, the trend-stationary hypothesis receives the highest posterior probability in all cases except for the Japanese yen/German mark real exchange rate. In that case, the stationary hypothesis receives the highest posterior probability. The French franc/German mark real exchange rate also deserves mention because it is the only case where the trend-stationary and the stationary hypotheses receive approximately the same posterior probabilities. Interestingly, these are the same two real exchange rates that Cheung and Lai [26] have found to be well characterized by stationary or

---

[7] Other researchers, Papell [28], and Culver and Papell [29], for example, argued that the presence of a linear deterministic time trend is inconsistent with long-run PPP, however.
[8] We have also tried other methods of lag-length determination, e.g., Akaike's Information Creteria. While the lag lengths determined tend to be much shorter, they do not produce significantly different results than those reported in Tables **1** and **2**, however.
[9] We use a non-Bayesian method, since the number of hypotheses increases rather rapidly if the same Bayesian approach is used to determine the lag lengths. This reduces the number of alternative hypotheses we need to consider and allow us to allocate the entire prior probability to the three hypotheses in question which are what we are most interested in. To check the robustness of our Bayesian results, we have also used a uniform four

lags for all the real exchange rate series. The results are not very different from those reported in the paper.
[10] Our later simulation results also suggest that the results in Table **2** are unlikely the result of the choice of the lag-length.

trend-stationary processes using the conventional ADF tests. The unit root hypothesis, on the other hand, receives no significant posterior probability. In sum, the Bayesian results strongly support the hypothesis that the real exchange rates are trend-stationary AR processes. Or, put differently, the restrictions on the null model are not strongly supported by the data. Thus, according to the "objective" Bayesian unit-root test, the results support long-run PPP.

Table 2.   Posterior Probabilities

Real Exchange Rate	k	Trend Stationary	Stationary	Unit Root
Canada/U.S.	11	0.83	0.14	0.03
U.K./U.S.	2	0.84	0.06	0.10
Germany/U.S.	2	0.91	0.06	0.03
Italy/U.S.	2	0.91	0.05	0.04
Japan/U.S.	2	0.85	0.11	0.04
France/U.S.	2	0.93	0.04	0.02
U.K./Canada	2	0.74	0.20	0.06
Germany/Canada	2	0.92	0.04	0.03
Italy/Canada	2	0.53	0.29	0.17
Japan/Canada	2	0.70	0.25	0.05
France/Canada	2	0.89	0.07	0.04
Germany/U.K.	2	0.80	0.15	0.05
Italy/U.K.	2	0.87	0.04	0.09
Japan/U.K.	3	0.89	0.07	0.04
France/U.K.	2	0.83	0.09	0.08
Italy/Germany	2	0.83	0.14	0.04
Japan/Germany	2	0.21	0.65	0.15
France/Germany	2	0.39	0.36	0.25
Japan/Italy	2	0.63	0.18	0.18
France/Italy	2	0.91	0.07	0.02
France/Japan	3	0.65	0.24	0.11

Note: The posterior probabilities may not sum to one because of rounding.

The two empirical approaches present rather different conclusions concerning long run PPP. One way to resolve this is to assess the reliability of each empirical approach. There is already a body of literature that suggests that classical unit-root tests such as the ADF tests have low power against plausible alternatives such as trend-stationary alternative (e.g., Hakkio [34], DeJong, Nankervis, Savin, and Whiteman [13]). Not much, however, is known about the power of the "objective" Bayesian unit-root test. As we mentioned earlier, Koop [1] presented some limited simulation results that suggest that the "objective" Bayesian unit-root test has fairly good power. We have also mentioned, however, that the test statistics used in Koop's simulations are dependent on the initial starting value of the DGM. Thus, Koop's simulation results are rather limited in its usefulness, and the general reliability of the "objective"

Bayesian unit-root test cannot be said to have been established. In the next section, we will investigate the reliability of the "objective" Bayesian results by performing a number of simulations that are more general that those presented by Koop [1].

4. THE POWER OF THE "OBJECTIVE" BAYESIAN UNIT-ROOT TEST

In this section, we perform a number of Monte Carlo experiments to study the power of the "objective" Bayesian unit-root test. This will allow us to assess the reliability of our results in Section 3, investigate the usefulness the "objective" Bayesian approach in empirical application, and whether or not it is better than the classical ADF approach in unit-root tests. The power of the "objective" Bayesian unit-root test in these experiments is defined as the percentage of time that the hypothesis corresponding to the DGM receives the highest posterior probability in repeated samples. The greater the percentage, the greater the power. For all our experiments, we use a DGM of the following form:

$$y_t = \delta + \rho y_{t-1} + \lambda t + u_t, \quad u_t \sim iid\ N(0, \sigma_u^2),$$

where $\underline{\theta} = (\delta, \rho, \lambda)$ is a vector of coefficients. Since the test statistics of our DGMs depend only on ρ, for the first experiment, we use $\underline{\theta} = (0.059317, 0.95 \le \rho \le 0.99, -0.0000201)$, $u_t \sim iid\ N(0, 0.000564)$, and ρ varies from 0.95 to 0.99, in increment of 0.01. Thus, the data generating model is a trend-stationary AR(1) process. The constant, the time-trend coefficient, and the variance of the error process are averages from estimating the null model (Equation 1) with data from the twenty-one real exchange rate series. Thus, they represent economically plausible parameters. The values of ρ that we choose are also based on what we consider to be plausible alternatives for monthly series. For example, Sims [2] argued that for monthly data, it is reasonable to concentrate the prior odds on the interval (0.98, 1) as opposed to, say (0.5, 1). Second, there appears to be a consensus among economists that deviations from long-run PPP have a half-life of about three to five years (see e.g., Abuaf and Jorion [35], Rogoff [22]).[11] According to Caner and Kilian [36], this corresponds to a ρ value of 0.98 and 0.99, respectively, for a half-life of three and five years, using monthly data. In sum, we believe that the parameter values that we choose for our Monte Carlo experiment are economically plausible.

For all our simulations, we first generate $T + 100$ observations, where T is the actual sample size used in our simulations, and discard the first 100 observations to avoid the initialization problem. All experiments are replicated 5,000 times for each ρ value. We start our simulation by letting $T = 311$, which corresponds to the sample size of our real exchange rates. The results are reported in the upper panel of Table **3**. The results reveal a

---

[11] A recent paper by Murray and Papell [21] has shown that the half-life estimates are extremely unreliable, however.

surprising pattern. For $0.95 < \rho < 0.98$, the trend-stationary hypothesis receives the highest posterior probability the greatest proportion of the time. The power of the test increases from $\rho = 0.95$ and peaks at $\rho = 0.97$. At $\rho = 0.99$, however, the unit-root hypothesis receives the highest posterior probability the largest proportion of the time, suggesting an extremely low power of the "objective" Bayesian test at this value of ρ.

Table 3.   (a) Data Generating Model: $y_t = 0.059317 + \rho y_{t-1} - 0.00000201t + u_t$, $u_t \sim iid\ N(0, 0.000564)$

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
0.99	3.64	0.02	96.34
0.98	68.18	16.36	15.46
0.97	73.00	14.80	12.20
0.96	66.76	10.30	22.94
0.95	55.44	6.16	38.40

(b) Data Generating Model: $y_t = 0.059317 + \rho y_{t-1} + u_t$, $u_t \sim iid\ N(0, 0.000564)$

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
1.00	67.22	32.32	0.46
0.99	5.08	0.10	94.82
0.98	75.26	3.84	20.90
0.97	79.76	2.16	18.08
0.96	69.60	1.36	29.04
0.95	56.34	0.52	43.13

We perform a second experiment using a DGM model with $\underline{\theta} = (0.059317, 0.95 \le \rho \le 1.00, 0.0)$, i.e., an AR(1) model, and u_t has the same properties as the first experiment. In this experiment, however, we vary ρ from 0.95 to 1.00, in increment of 0.01. Thus, the DGM is a stationary AR(1) model for $0.95 < \rho < 0.99$, and a random-walk with drift model when $\rho = 1$. The results are reported in the lower panel of Table **3**. Interestingly, for $0.95 < \rho < 0.99$, the results mirror the results reported in the upper panel of Table **3**. This, perhaps, is not surprising since the effect of the time trend is extremely small, thus making the two DGMs almost the same. At $\rho = 1.00$, the trend-stationary hypothesis receives the highest posterior probability the largest percentage of the time, followed by the stationary

hypothesis, while the unit-root hypothesis receives the highest posterior probability less than 1% of the time.

The results in Table **3** are not very encouraging to the "objective" Bayesian unit-root test. First, for $0.95 \le \rho \le 0.98$, it cannot distinguish between a highly persistent trend-stationary model from a highly persistent stationary AR model. There appears to be a bias in favor of the trend-stationary model. At $\rho = 0.99$ the unit-root hypothesis is favored regardless of the DGM used. When the DGM is a random-walk with drift model, the "objective" Bayesian unit-root is biased in favor of the trend-stationary model. Thus, just as the classical ADF test is criticized frequently for its bias in favor of finding an unit root, it appears that the "objective" Bayesian unit-root test can also be criticized for its bias in favoring trend-stationarity. Given the Monte Carlo results, it is impossible to draw any conclusions regarding long-run PPP since the underlying DGMs of the real exchange rates are unknown. Thus, the simulation results suggest that, because of its low power against plausible alternatives, the "objective" Bayesian unit-root test does not provide a better statistical approach than the classical ADF test in unit-root testing.

We also perform a number of additional Monte Carlo experiments to determine if the "objective" Bayesian unit-root test may be sensitive to sample size, the size of the constant, or the size of the trend coefficient. We use the same two DGMs as earlier but with $u_t \sim iid\ N(0, 1)$, and again generate $T + 100$ observations for the simulations. In particular, we use sample sizes of $T = 60, 240, 720$, which are approximately the sample sizes of post-war annual, quarterly, and monthly data, respectively. For the constant term, we use $\delta = 1, 0$, and -1, and for the trend-stationary model, we use $\lambda = 0.05$, and -0.05. We perform simulations for all possible combinations of T, δ, and λ, using both the AR(1) and the trend-stationary AR(1) DGMs. All experiments are also replicated 5,000 times.

For both the trend-stationary AR(1) and the AR(1) DGMs, the results are not very sensitive to the constant and the trend coefficient that we use but are sensitive to the sample size. Rather than reporting the large number of results, we will instead report in Tables **4** and **5** only the representative results. Table **4** shows the results for the trend-stationary AR(1) DGM with $\underline{\theta} = (1.00, 0.95 \le \rho \le 0.99, 0.05)$. For $T = 60$, the "objective" Bayesian unit-root test cannot easily distinguish between the trend stationary and the stationary hypotheses. Its power, however, increases gradually as ρ approaches 0.99. For the two larger samples, the power of the "objective" test drops off significantly. That is, the proportion of times that the hypothesis that corresponds to the DGM receiving the largest posterior probability actually decreases as the sample size increases. This is in contrast to classical statistics, where generally the power of a test increases as the sample size increases as the bias due to small sample diminishes. For $0.95 \le \rho \le 0.98$, the stationary hypothesis receives the largest posterior probability the largest percentage of the time. For $\rho = 0.99$,

the "objective" test favors the unit-root hypothesis almost 100 percent of the time. Thus, for the two larger samples, there is a bias in favor of the stationary hypothesis, except when $\rho = 0.99$ where the unit-root hypothesis is favored. In summary, the results in Table **4** show that the "objective" Bayesian unit-root test has very low power when the DGM is a trend-stationary process.

Table 4.   Data Generating Model: $\underline{\theta}$ = (1.00,0.95≤ρ≤0.99,0.05), u_t ~ iid N(0,1)

T = 60

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
0.99	51.00	33.40	15.60
0.98	49.52	48.68	1.80
0.97	48.06	50.90	1.04
0.96	45.40	53.52	1.08
0.95	43.98	55.10	0.92

T = 240

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
0.99	0.10	0.00	99.90
0.98	49.02	50.72	0.26
0.97	38.74	61.24	0.02
0.96	26.26	73.74	0.00
0.95	16.10	83.90	0.00

T = 720

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability:		
ρ	Trend Stationary	Stationary	Unit Root
0.99	0.00	0.00	100.00
0.98	13.64	86.36	0.00
0.97	2.04	97.96	0.00
0.96	0.06	99.94	0.00
0.95	0.00	100.00	0.00

Turning now to Table **5** where we use an AR(1) DGM with $\underline{\theta} = (1.00, 0.95 \leq \rho \leq 1.00)$, we see that the power of the "objective" Bayesian test is extremely low. For $T = 60$ and $T = 240$, there is a clear bias in favor of the trend stationary hypothesis, and also for $T = 720$ when $0.98 \leq \rho \leq 1.00$. The unit root hypothesis is favored

when $0.95 \leq \rho \leq 0.97$ for $T = 720$. Again, just as the results in Table **4**, the power of the "objective" Bayesian unit-root test also decreases as the sample size increases.

Table 5.   Data Generating Model: $\underline{\theta}$ = (1.00,0.95≤ρ≤1.00), u_t ~ iid N(0,1)

T = 60

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
1.00	51.00	46.66	2.34
0.99	59.94	32.44	7.62
0.98	69.16	20.52	10.32
0.97	70.88	17.22	11.90
0.96	70.62	15.30	14.08
0.95	71.42	13.66	14.92

T = 240

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
1.00	64.80	34.58	0.62
0.99	75.02	11.24	13.74
0.98	82.92	7.22	9.86
0.97	81.78	4.30	13.92
0.96	76.80	3.58	19.62
0.95	68.22	2.08	29.70

T = 720

	Proportion of Times that a Hypothesis Receives the Largest Posterior Probability		
ρ	Trend Stationary	Stationary	Unit Root
1.00	74.92	24.86	0.22
0.99	63.38	1.16	35.46
0.98	71.60	1.34	27.02
0.97	43.58	4.20	52.22
0.96	15.00	0.00	85.00
0.95	2.60	0.00	97.40

Finally, we are also interested in knowing whether or not the power of the "objective" Bayesian test would change if the DGM is less persistent. We repeat our simulations for the three sample sizes with the AR(1) DGM, for $\delta = 1.00$, and $\rho = 0.70, 0.75, 0.80, 0.85, 0.90$. The results are shown in Table **6**. Once again, the power of the "objective" Bayesian

test is extremely low. For the smallest sample size, the "objective" Bayesian test favors either the trend-stationary hypothesis or the unit-root hypothesis. For the two larger sample sizes, the unit-root hypothesis is favored most of the time. We again observe that the power of the "objective" Bayesian unit-root test decreases with the sample size.

Table 6. **Data Generating Model: $\underline{\theta}$ = (1.00, ρ), $u_t \sim iid\ N(0,1)$**

T = 60

Proportion of Times that a Hypothesis Receives the Largest Posterior Probability			
ρ	Trend Stationary	Stationary	Unit Root
0.90	66.38	6.94	26.68
0.85	55.70	3.24	41.06
0.80	40.68	2.22	57.10
0.75	24.28	2.20	73.52
0.70	12.92	0.04	87.04

T = 240

Proportion of Times that a Hypothesis Receives the Largest Posterior Probability			
ρ	Trend Stationary	Stationary	Unit Root
0.90	14.18	0.04	85.78
0.85	0.46	0.00	99.54
0.80	0.00	0.00	100.00
0.75	0.00	0.00	100.00
0.70	0.00	0.00	100.00

T = 720

Proportion of Times that a Hypothesis Receives the Largest Posterior Probability			
ρ	Trend Stationary	Stationary	Unit Root
0.90	0.00	0.00	100.00
0.85	0.00	0.00	100.00
0.80	0.00	0.00	100.00
0.75	0.00	0.00	100.00
0.70	0.00	0.00	100.00

The simulation results in this section lead us to several conclusions. First, regardless of the DGMs used, the "objective" Bayesian unit-root test has rather low power. Moreover, the general tendency is to favor the trend-stationary hypothesis. Second, the test statistics are sensitive to the sample size. Contrary to what one would expect that the power of the test to increase with the sample size as the small sample bias diminishes, the power of the "objective" Bayesian unit-root test actually decreases with the sample

size increases. Thus, our simulation results show that the "objective" Bayesian unit-root test is rather unreliable and certainly does not provide an improvement over the classical approach in unit-root test.

5. SUMMARY AND CONCLUSIONS

Researchers generally agree that the Bayesian approach offers a useful alternative to the classical approach in empirical modeling. In unit-root testing, Sims [2], Sims and Uhlig [8], and Koop [1, 10] have advocated the Bayesian approach over the classical ADF tests. Phillips [3, 4] identified the need for subjective priors as one of the criticisms of the Bayesian approach, and called for more objective Bayesian analysis. Koop [1] offered the "objective" Bayesian unit-root test which requires less subjective input from the researcher, is computationally simple, and thus offers an attractive alternative to the classical approach in unit-root testing. Koop [1] offered some limited simulation results on the size and power of the "objective" Bayesian unit-root test, and found them to be mostly favorable. Except for Koop's [1] paper, by and large, the power of the "objective" Bayesian unit-root test is unknown. In this paper, we study the power of Koop's "objective" unit-root test. We use a general DGM, conduct the simulations using different sample sizes and different parameter values. Our objective is to provide more evidence on the reliability of the "objective" Bayesian unit-root test. In particular, we are interested in whether or not it provides a better alternative to the classical ADF unit-root test, and whether or not the use of "objective" priors is appropriate.

We first test for long-run PPP using both the classical ADF and the "objective" Bayesian tests. The results with the classical ADF offer little support for long-run PPP. Our "objective" Bayesian results, on the other hand, provide a stark contrast to the ADF results. In all cases, the hypothesis of an unit root does not receive significant posterior probability. Rather, sample information appears to strongly support the hypothesis of trend-stationarity for all cases except the Japanese yen/German mark real exchange rate where the sample information suggests a stationary time series. The French franc/German mark real exchange rate is the other case where the stationary hypothesis receives significant posterior probability.

Next, we study the power of the "objective" Bayesian test with Monte Carlo simulations first using parameter values suggested by the real exchange rate sample data and then using more general parameter values. The results are not very encouraging. In particular, using economically plausible parameters for monthly data of real exchange rates for our DGMs, we find that the "objective" Bayesian test cannot distinguish between a trend-stationary AR model from a stationary AR model when the time trend effect is relatively small, and the time series is highly persistence. The bias is in favor of finding a trend-stationary model. When $\rho = 0.99$, the "objective" Bayesian test is biased in favor of a unit-root. On the other hand, when the DGM is a random-walk with drift model, the test is biased in favor of the trend-stationary hypothesis. Additional simulations also

suggest that the "objective" Bayesian test is sensitive to sample sizes, where the power actually decreases as the sample size increases.

In summary, contrary to Koop's [1] findings, our simulation results suggest that the "objective" Bayesian unit-root test is not a reliable test since there is a bias in favor of finding a trend-stationary model. Our application to a test of long-run PPP clearly demonstrated the danger of using the "objective" Bayesian unit-root test in empirical testing. Thus, we conclude that there is no evidence to suggest that the "objective" Bayesian test is better than the classical ADF tests in unit-root tests. We see nothing here to recommend its use in unit-root testing. Thus, unfortunately, while there is a need for more objective analysis of Bayesian time series, Koop's "objective" Bayesian test does not appear to move us closer to that goal.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Koop G. 'Objective' Bayesian unit root tests. J Appl Econom 1992; 7: 65-82.
[2]   Sims C. Bayesian skepticism on unit root econometrics. J Econ Dyn Control 1988; 12: 463-74.
[3]   Phillips P. To criticize the critics: an objective analysis of stochastic trends. J Appl Econom 1991; 6: 333-64.
[4]   Phillips P. Bayesian unit roots: de rebus priorbus simper est disputandum. J Appl Econom 1991; 6: 435-73.
[5]   Nelson C, Plosser C. Trends and random walks in macroeconomics time series: some evidence and implications. J Monet Econ 1982; 10: 139-62.
[6]   Murray C, Nelson C. The uncertain trend in U.S. GDP. J Monet Econ 2000; 46: 79-95.
[7]   DeJong D, Whiteman C. The case for trend-stationarity is stronger than we thought. J Appl Econom 1991; 6: 413-21.
[8]   Sims C, Uhlig H. Understanding unit rooters: a helicopter tour. Econometrica 1991; 59: 1591-9.
[9]   Dickey D, Fuller W. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 1981; 49: 1057-72.
[10]   Koop G. Recent progress in applied Bayesian econometrics. J Econ Surv 1994; 8: 1-34.
[11]   Pesaran MH, Ed. New York: John Wiley & Sons. J Appl Econom 1991; Vol. 6(4).
[12]   Zellner A, Siow A. Posterior odds for selected regressions. In: Bernardo J, DeGroot M, Lindley D, Smith A Eds. Bayesian Statistics, Valencia: University Press 1980; pp. 583-603.
[13]   DeJong D, Nankervis J, Savin N, Whiteman C. Integration versus trend stationarity in time series. Econometrica 1992; 60: 423-33.
[14]   Frenkel J. A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. In: Frenkel J, Johnson H, Eds. The economics of exchange rate. Reading, MA: Addison-Wesley 1978; pp. 1-25.
[15]   Dornbusch R. Expectations and exchange rate dynamics. J Pol Econ 1976; 84: 1161-76.
[16]   Sarno L, Taylor M. Purchasing power parity and the real exchange rate. IMF Staff Papers 2002; 49: 65-105.
[17]   Ahking F. Non-parametric tests of real exchange rate in the post-Bretton Woods era. 2009 (in press).
[18]   Diebold F, Husted S, Rush M. Real exchange rates under the gold standard. J Pol Econ 1991; 99: 1252-71.
[19]   Engel C, Kim CJ. The long-run U.S./U.K. real exchange rate. J Money Credit Banking 1999 Part 1; 31: 335-56.
[20]   O'Connell P. The overvaluation of purchasing power parity. J Inter Econ 1998; 44: 1-19.
[21]   Murray C, Papell D. The purchasing power parity persistence paradigm. J Inter Econ 2002; 56: 1-19.
[22]   Rogoff K. The purchasing power parity puzzle. J Econ Lit 1996; 34: 647-68.
[23]   Schotman P, Van Dijk H. A Bayesian analysis of the unit root in real exchange rates. J Econom 1991; 49: 195-238.
[24]   Whitt J. The long-run behavior of the real exchange rate. J Money Credit Bank 1992; 24: 72-82.
[25]   Ahking F. Testing long-run purchasing power parity with a Bayesian unit root approach: the experience of Canada in the 1950s. Appl Econ 1997; 29: 813-19.
[26]   Cheung YW, Lai KS. Parity reversion in real exchange rates during the post-Bretton Woods period. J Inter Money Financ 1998; 17: 597-614.
[27]   Koedijk K, Schotman P, Van Dijk M. The re-emergence of PPP in the 1990s. J Inter Money Fin 1998; 17: 51-61.
[28]   Papell D. Searching for stationarity: purchasing power parity under the current float. J Inter Econ 1997; 43: 313-32.
[29]   Culver S, Papell D. Long-run purchasing power parity with short-run data: evidence with a null hypothesis of stationarity. J Inter Money Financ 1999; 18: 751-68.
[30]   Ng S, Perron P. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. J Am Stat Assoc 1995; 90: 268-81.
[31]   Perron P. Further evidence on breaking trend functions in macroeconomic variables. J Econom 1997; 80: 355-85.
[32]   Fuller W. Introduction to statistical time series. New York, New York: John Wiley & Sons 1976.
[33]   Cheung YW, Lai KS. Lag order and critical values of the augmented Dickey-Fuller test. J Bus Econ Stat 1995; 13: 277-80.
[34]   Hakkio C. Does the exchange rate follow a random walk? A Monte Carlo study of four tests for a random walk. J Inter Money Financ 1986; 5: 221-9.
[35]   Abuaf N, Jorion P. Purchasing power parity in the long run. J Financ 1990; 45: 157-74.
[36]   Caner M, Kilian L. Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. J Inter Money Fin 2001; 20: 639-57.