Journal of Econometrics 49 (1991) 195–238. North-Holland

A Bayesian analysis of the unit root in real exchange rates

Peter Schotman and Herman K. van Dijk*
Erasmus University, 3000 DR Rotterdam, The Netherlands

We propose a posterior odds analysis of the hypothesis of a unit root in real exchange rates. From a Bayesian viewpoint the random walk hypothesis for real exchange rates is *a posteriori* as probable as a stationary AR(1) process for four out of eight time series investigated. The French franc/German mark is clearly stationary, while the Japanese yen/US dollar is most likely a random walk. In contrast, classical tests are unable to reject the unit root for any of these series.

1. Introduction

Nominal and real exchange rates behave almost like random walks. This conclusion emerges from much of the recent empirical literature on exchange rate models. In a series of papers Meese and Rogoff (1983a, b, 1988) compared the forecasting performance of many econometric models, including structural models, unrestricted VARs, and univariate time series models, and found that none outperformed a simple random walk. Similar results have been obtained in other studies. Frankel and Meese (1987) and Dornbusch and Frankel (1987) review the evidence and its implications. Especially the random walk results for *real* exchange rates have serious consequences for economic theories of exchange rate behaviour. If the real exchange rate follows a random walk, shocks to the real exchange rate accumulate and a time series of real exchange rates will not show a tendency of mean reversion. This is contrary to the notion of Purchasing Power Parity (PPP) which posits that there is a constant equilibrium real exchange rate. In

*We thank seminar participants at CORE (Louvain-la-Neuve), Cornell University, London School of Economics, ESEM (Munich), and the SBIES meeting (Chicago) for stimulating discussions. John Geweke, Nick Kiefer, Frank Kleibergen, Dale Poirier, Arnold Zellner, and two anonymous referees provided helpful comments. We thank Frank Kleibergen for computational assistance. Errors are of course our own responsibility. The second author acknowledges financial support from NWO.

most monetary type exchange rate models [see Dornbusch (1976) and Frankel (1979)] the constant PPP level functions as an anchor to which prices and nominal exchange rates continually adjust.

The forecasting experiments are not the only empirical evidence that suggest a random walk model for real exchange rates. Formal statistical tests do not reject the null hypothesis of a unit root against the alternative of a stationary autoregressive time series model. This conclusion was reached by several authors using different test functions. Meese and Rogoff (1988) tested against a unit root in monthly data for the US dollar vis-á-vis the German mark, the Japanese yen, and the British pound during the floating period and could not reject the unit root hypothesis using the standard Dickey–Fuller test and the more robust Phillips (1987) test. Edison and Fisher (1988) reach the same conclusion for most currencies within the EMS with the exception of the German mark against the Dutch guilder. Huizinga (1987) and Kaminsky (1988) use the variance ratio developed by Cochrane (1988) to test the random walk against more general time series models for real exchange rates. While these authors find some tendencies of mean reversion, they cannot formally reject the null hypothesis of a random walk. Diebold (1988) and Mecagni and Pauly (1987) use both consumer prices as well as wholesale prices to construct real exchange rates with the idea that wholesale prices better reflect the prices of tradeable goods that are appropriate for tests of long-run PPP. Again their univariate tests cannot reject the unit root hypothesis, suggesting that the measurement of the macroeconomic price index is probably not the most critical assumption.

As all these authors admit, part of the explanation for these results is the low power of unit root tests and the relatively short sample period of floating exchange rates, starting in 1973. This means that we have at most about 190 monthly observations. Even in the ideal circumstance that the true model is an AR(1), a classical statistic like the first-order autocorrelation must be less than 0.92 in order to reject the unit root.[1] The observed autocorrelations in actual real exchange rate series are 'consistent' with a random walk null hypothesis. That does of course not imply that the random walk is the most likely description of the time series process of real exchange rates. A stationary alternative with a first-order autocorrelation parameter of 0.97 might explain the data equally well.

A second problem with the application of unit root tests to real exchange rates is that the null hypothesis in these tests is that the theory is false. The random walk model for real exchange rates implies violation of long-run PPP. This is contrary to the methodology of testing in the sense that testing an economically interesting hypothesis at the 5% level means that there is a 5%

[1]The 5% critical value of the test statistic $n(\hat{\rho}-1)$ in Fuller (1976) is approximately -13.9 for the 190 observations that are available in the floating rate period.

chance to reject the economic theory while it is true; true here means that the time series is stationary.[2] Classical hypothesis testing must have the random walk as the null hypothesis instead of a stationary AR(1); the random walk comprises a particular subset of the class of AR(1) models. In a Bayesian approach the null and alternative hypothesis can be treated symmetrically. There is no need to look at the data from the specific viewpoint of stationarity or nonstationarity. Given the data one can determine which of the two is the most likely. The formal Bayesian tool for choosing between different models is the computation of posterior odds; see, e.g., Leamer (1978, sec 4.3.) or Zellner (1971, pp. 297–298). By assigning a discrete probability to the occurrence of a random walk posterior odds can be used to test a sharp null hypothesis. In this paper we will develop a posterior odds ratio for choosing between a random walk and a stationary AR(1) model. The purpose of our study is to reexamine the random walk results for real exchange rates. Is the random walk still the most favoured model if compared directly to a simple plausible alternative?

Some first results from a Bayesian perspective are provided by Sims (1988) and DeJong and Whiteman (1989). Sims (1988) considers the AR(1) model without a constant term. Sims' test procedure relies on computation of a posterior odds ratio for the unit root hypothesis versus a stationary alternative. In this paper we will extend Sims' results by introducing an unknown constant term in the model. It will be shown that with an uninformative prior on the mean of the process the posterior odds test breaks down, since the posterior density of the first-order autocorrelation coefficient is improper. This result derives from the fact that the mean of a random walk does not exist. The aim of the paper is to provide a test procedure that makes efficient use of the information in the data. In specifying prior distributions we will take care to avoid unreasonable formulations of the prior that dominate the posterior odds, but are not supported by the data.

We will focus our analysis of unit roots on the simple case of a univariate first-order autoregressive process, as treated in Fuller (1976). Even for this simple process the literature distinguishes between three different specifications: models without a constant term, with a constant term, and with constant term and linear time trend. This distinction is useful, since one of the difficulties with classical tests is that the distribution of the test statistic depends on the presence of nuisance parameters like a constant and a trend [see Evans and Savin (1984)]. Diagnostic screening of the data (see section 4) suggests that this class of models is also rich enough for the application to real exchange rates.

[2]Christiano and Eichenbaum (1989) and Campbell and Mankiw (1987) provide a different approach to unit root tests. They start off with taking first differences of a time series and then test whether this has led to overdifferencing. Their null hypothesis is stationarity and the alternative is a unit root.

The organization of the paper is as follows. First, in sections 2 and 3, we derive some properties of posterior odds for unit root hypotheses in AR(1) models. Second, in section 4, we apply this tool to tests of long-run PPP for eight different real exchange rates. Section 5 concludes with some final remarks.

2. The AR(1) model without constant term

2.1. Definitions and model specification

In order to concentrate on the differences between the classical unit root tests and the Bayesian procedure we start off with the simplest possible model, a first-order autoregressive process with mean zero. Suppose that we have a sample of T consecutive observations on a time series $\{y_t\}$ generated by

$$y_t = \rho y_{t-1} + u_t, \tag{1}$$

where we further assume that:

(i) y_0 is a known constant,
(ii) u_t are identically and independently (i.i.d.) normally distributed with mean zero and unknown variance σ^2,
(iii) $\rho \in S \cup \{1\}$; $S = \{\rho \mid -1 < a \leq \rho < 1\}$.

The econometric analysis aims at discriminating between a stationary model (here defined as $a \leq \rho < 1$) and the nonstationary model with $\rho = 1$, which we will henceforth refer to as the random walk (RW)[3]. Assumption (i) implies that we will work conditional on initial observations. The reason for making this assumption is that treatment of the initial conditions will differ between the stationary and the nonstationary case, so that the specification of a simple marginal distribution for y_0 is a nontrivial problem; see Sims (1988, sec. 6). The lower bound a in assumption (iii) largely determines the specification of the prior for ρ.

The principal Bayesian tool to compare a sharp null hypothesis with a composite alternative hypothesis is the posterior odds ratio [see Zellner

[3]More general definitions of stationarity are given in, e.g., Spanos (1986, ch. 8). We will frequently use the terms stationary and nonstationary instead of the more appropriate terminology of 'integrated of order zero' [$I(0)$] and 'integrated of order one' [$I(1)$]. Within the class of autoregressive processes a time series is called $I(0)$ if its characteristic function has all roots strictly outside the unit circle. The series is called $I(1)$ if one of the roots is unity. In the autoregressive case the terms stationary for $I(0)$ and nonstationary for $I(1)$ will be used as synonyms, although the absence of a unit root does not automatically imply stationarity of the series, except in the AR(1) case.

(1971, ch. 10)]. For our case the odds ratio is defined as

$$K_1 = K_0 \frac{\int_0^\infty p(\sigma) L(y|\rho=1,\sigma,y_0)\,\mathrm{d}\sigma}{\int_S \int_0^\infty p(\sigma)p(\rho)L(y|\rho,\sigma,y_0)\,\mathrm{d}\sigma\,\mathrm{d}\rho} = \frac{\Pr(\rho=1|Y)}{\Pr(\rho\in S|Y)}, \qquad (2)$$

where

K_0 = prior odds in favour of the hypothesis $\rho = 1$,
K_1 = posterior odds in favour of the hypothesis $\rho = 1$,
$p(\rho)$ = prior density of $\rho \in S$,
$p(\sigma)$ = prior density of σ,
$L(y|\cdot)$ = likelihood function of the observed data $y = (y_1, \ldots, y_T)'$,
Y = $(y_0\ y')'$, all observed data.

The posterior odds K_1 are equal to the prior odds K_0 times the Bayes factor. The Bayes factor is the ratio of the marginal posterior density of ρ under the null hypothesis $\rho = 1$ to a weighted average of the marginal posterior under the alternative using the prior density of ρ as a weight function. The prior odds express the special weight given to the null hypothesis; the point $\rho = 1$ is given the discrete prior probability $\vartheta = K_0/(1+K_0)$. From the posterior odds one can compute the posterior probability of the null hypothesis as $K_1/(1+K_1)$.

For the complete specification of the marginal prior of ρ and σ we assume that

$$\Pr(\rho = 1) = \vartheta, \qquad (3)$$

$$p(\rho|\rho \in S) = 1/(1-a), \qquad (4)$$

$$p(\sigma) \propto 1/\sigma. \qquad (5)$$

The prior of ρ is uniform on S but has a discrete probability ϑ that $\rho = 1$.[4] The prior on σ is diffuse, and corresponds to a uniform prior on $\ln\sigma$. We assume that ρ and σ are independent.

[4]One alternative would be a Beta prior on ρ; see, e.g., Zellner (1971, ch. 7). The selection of the parameters of the Beta prior is not easy. Sensible choices will, a.o., depend on the observation frequency of the data. Monthly data require a prior that is more concentrated to the unit root than annual data. This will also affect the choice of the lower bound a; see below.

The likelihood function for the vector of T observations y reads

$$L(y|\rho,\sigma,y_0)=(2\pi\sigma^2)^{-T/2}\exp\left(-\frac{1}{2\sigma^2}u'u\right), \tag{6}$$

where $u=(y-y_{-1}\rho)$ and $y_{-1}=(y_0,\dots,y_{T-1})'$. Combining likelihood and prior yields the posterior distribution of the parameters.

2.2. Computation of posterior odds

To compute the posterior odds we integrate the posterior density over the nuisance parameter σ using the integration formula

$$\int_0^\infty \sigma^{-(T+1)}\exp\left(-\frac{1}{2\sigma^2}u'u\right)\mathrm{d}\sigma=\tfrac{1}{2}\Gamma(T/2)\left(\tfrac{1}{2}u'u\right)^{-T/2}. \tag{7}$$

The right-hand side of (7) can be rewritten as the kernel of a Student-t density. The numerator of the posterior odds ratio is simply the value of the t-density evaluated at $\rho=1$. For the denominator we have to integrate over ρ. We make use of the properties of the t-distribution to obtain

$$\begin{aligned}\int_a^1(u'u)^{-T/2}\mathrm{d}\rho&=\int_a^1\left((y'_{-1}y_{-1})(\rho-\hat\rho)^2+(T-1)\hat\sigma^2\right)^{-T/2}\mathrm{d}\rho\\&=C_T\left((T-1)\hat\sigma^2\right)^{-T/2}\left((T-1)s_{\hat\rho}^2\right)^{1/2}\\&\quad\times\left(F\left(\frac{1-\hat\rho}{s_{\hat\rho}};T-1\right)-F\left(\frac{a-\hat\rho}{s_{\hat\rho}};T-1\right)\right),\end{aligned} \tag{8}$$

where

$\hat\rho=\dfrac{y'_{-1}y}{y'_{-1}y_{-1}}$, the OLS estimator of ρ,

$s_{\hat\rho}^2=\hat\sigma^2(y'_{-1}y_{-1})^{-1}$, the squared OLS standard error of $\hat\rho$,

$\hat\sigma^2=\dfrac{1}{T-1}\left(y'y-\dfrac{(y'_{-1}y)^2}{(y'_{-1}y_{-1})}\right)$, the estimated variance of the residuals,

$C_T=\dfrac{\Gamma((T-1)/2)\Gamma(1/2)}{\Gamma(T/2)}$, with $\Gamma(\cdot)$ the Gamma function,

and $F(x;\nu)$ is the standard cumulative t-density with ν degrees of freedom evaluated at x.

Having computed the relevant integrals the posterior odds ratio becomes

$$K_1 = \frac{C_T^{-1}}{(T-1)^{1/2}} \frac{\vartheta}{1-\vartheta} \left(\frac{\sigma_0^2}{\hat{\sigma}^2} \right)^{-T/2} \frac{1-a}{s_{\hat{\rho}}}$$

$$\times \left[F\left(\frac{1-\hat{\rho}}{s_{\hat{\rho}}}; T-1 \right) - F\left(\frac{a-\hat{\rho}}{s_{\hat{\rho}}}; T-1 \right) \right]^{-1}, \qquad (9)$$

where σ_0^2 is the variance of the first difference Δy_t, here defined as

$$\sigma_0^2 = \frac{1}{T-1} (y - y_{-1})' (y - y_{-1}).$$

The posterior odds ratio K_1 consists of four factors:

(i) the prior odds $\vartheta/(1-\vartheta)$,
(ii) the likelihood ratio $(\hat{\sigma}^2/\sigma_0^2)^{-T/2}$,
(iii) the length of the prior interval $(1-a)$ in units of $s_{\hat{\rho}}$, the scale parameter of the posterior,
(iv) the area under a truncated t-distribution.

The last three factors together form the Bayes factor. The expression for the odds in (9) is analogous to the posterior odds ratio for comparing hypotheses within a linear regression model. For an extensive discussion of the interpretation of (9), and how it can (or should) be used for Bayesian inference, we therefore refer to standard textbooks, like Zellner (1971, ch. 10).

To convert a posterior odds analysis based on (9) into an operational test procedure one should make sensible choices for a and ϑ. As in Sims (1988) the prior odds K_0 are taken to be balanced between the random walk and the stationary AR(1), i.e., we take $\vartheta = \frac{1}{2}$. Our specification of the lower bound a differs, however. In Sims (1988) a is a function of the observation frequency of the time series. This partly reflects his concern with having a prior that is concentrated in the range of ρ values for which the likelihood is large. Our determination of a is one possible formalization of this concern. It will be data-based; within the context of this model we have been unable to specify a genuine prior.

The length of the interval $[a, 1)$ directly enters the posterior odds ratio (9). If a is set at a very small value there will be a large interval where the posterior density of ρ will have almost no mass. Hence the 'averaged' likelihood in the denominator of the posterior odds ratio becomes small, since the average is taken over an interval that includes a large region where the likelihood is close to zero. If, for example, $a = 0$ or even $a = -1$ and the data indicate that the posterior is concentrated near $\rho = 1$, one will almost always reject the stationary model and accept the unit root. On the other hand one should also avoid taking a too close to unity. In that situation the prior excludes a region of ρ where the likelihood function has a nonnegligible mass.[5] The sensitivity of our empirical test results with respect to the choice of a will be discussed in section 4.

We choose a locally uniform prior on ρ on $[a, 1)$ and consider an empirical determination of the lower bound a. The constructed interval $[a, 1)$ contains a fraction $1-\alpha$ of all the probability mass of the posterior truncated t-distribution to the left of unity. This is the interval where the likelihood is 'large'. Reasonable values of α are between 0.001 and 0.1, so that the posterior captures between 90% and 99.9% of the total probability mass of ρ. The empirical lower bound a^* can then be expressed as

$$a^* = \hat{\rho} + s_{\hat{\rho}} F^{-1}(\alpha F(-\hat{\tau})), \tag{10}$$

where $F(\cdot)$ is the cumulative t-distribution, here with $T-1$ degrees of freedom, and $\hat{\tau} = (\hat{\rho} - 1)/s_{\hat{\rho}}$ is the Dickey–Fuller test statistic. Fig. 1 depicts the typical situation with $\alpha = 0.01$. It shows two posterior densities of ρ with $\hat{\rho} < 1$ and sample sizes T and $2T$. The lower bound a^* divides the interval $(0, 1)$ such that a fraction α of the total area (A) under the curve is to the left of a and a fraction $1-\alpha$ between a and unity. Thus a^* is a function of the data. Notice that the lower bound increases with sample size. The prior will be more concentrated as $T \to \infty$. In the posterior odds test we compare the unit root with ever smaller alternative regions. One consequence of the data-based choice of a^* is that Lindley's or Jeffreys' paradox does not apply to our empirical results.

After fixing numerical values for ϑ and α the posterior odds are just a function of the data like any other test statistic. Due to the specific way that the lower bound has been constructed, the posterior odds are directly related to the Dickey–Fuller test. Setting the prior odds equal to one, using (10) to substitute for a, and taking logarithms in (9) we get

$$\ln K_1 = c_T - \frac{T}{2}\ln\left(\sigma_0^2/\hat{\sigma}^2\right) + \ln\left(\frac{-\hat{\tau} - F^{-1}(\alpha F(-\hat{\tau}))}{F(-\hat{\tau})}\right), \tag{11}$$

where $c_T = -\frac{1}{2}\ln((T-1)C_T^2)$. For large T the constant c_T approaches $-\frac{1}{2}\ln(2\pi)$, the first term becomes $-\frac{1}{2}\hat{\tau}^2$, and $F(\cdot)$ tends to the cumulative

[5]This is a general problem in Bayesian testing of a point null hypothesis. See, e.g., Berger and Delampady (1987) and the references cited therein.

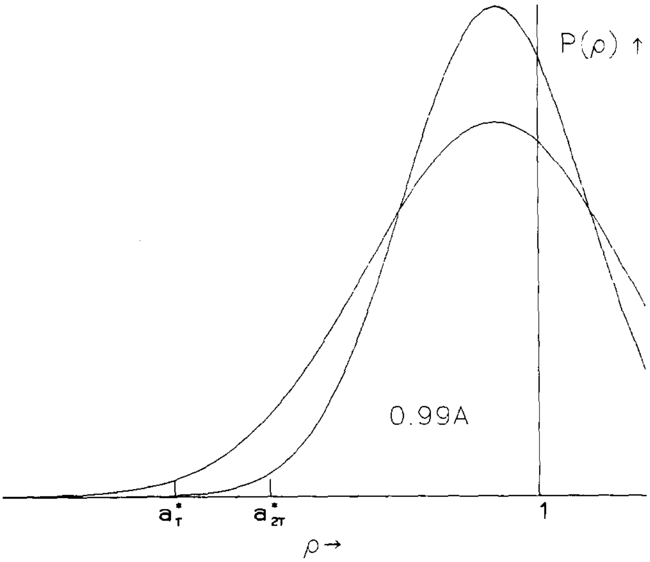


Fig. 1. Construction of lower bound for sample sizes T and $2T$.

normal distribution. The posterior odds then become a function of the Dickey–Fuller statistic $\hat{\tau}$:

$$\ln K_1 = -\tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}\hat{\tau}^2 + \ln\left(\frac{-\hat{\tau} - F^{-1}(\alpha F(-\hat{\tau}))}{F(-\hat{\tau})}\right). \tag{12}$$

Fig. 2 shows the relation between $\hat{\tau}$ and $\ln K_1$ for several values of α. The functions are monotonically increasing, and do not depend greatly on α. In the sequel we will always take $\alpha = 0.01$. For this value of α the 5% point of the Dickey–Fuller test under the null hypothesis of a random walk ($\hat{\tau} = -1.95$) coincides with a posterior odds ratio $K_1 = 0.27$, or equivalently a posterior probability $\Pr(RW) = 0.21$. According to the odds interpretation the classical Dickey–Fuller test rejects the random walk, if the posterior probability of the random walk is less than 0.21. For the 10% point of the Dickey–Fuller test the corresponding odds and posterior probability are 0.46 and 0.32 respectively. A posterior odds ratio of unity implies that $\hat{\tau} = -1.05$, approximately the 25% point of the Dickey–Fuller distribution under the null. In other words, the Dickey–Fuller test strongly favours the random walk, giving it a prior probability of 0.79. Since the posterior odds are a function of the Dickey–Fuller test statistic its sampling properties correspond exactly to those of the Dickey–Fuller test. From a classical point of view the posterior odds ratio is a test with a size of 0.25.

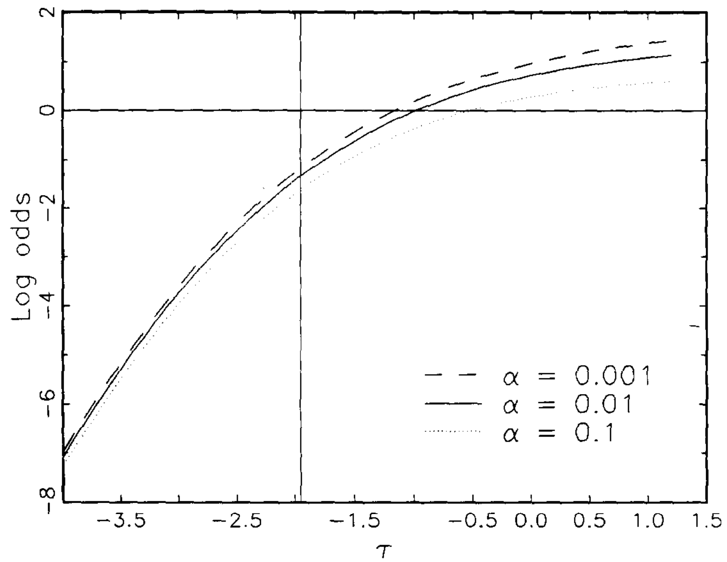


Fig. 2. Functional relation between posterior odds and Dickey–Fuller test.

3. The AR(1) model with a constant term

3.1. Representation

The presence of an unknown (unconditional) mean will complicate the Bayesian test procedure. The unconditional mean only enters the model in the stationary case. Under the unit root hypothesis it does not exist. Hence it is not a simple nuisance parameter like σ that can easily be integrated out. The model with an unknown mean can be written in several representations. The most direct way is to add a constant to the specification of section 2:

$$y_t = x_t + \mu, \tag{13a}$$

$$x_t = \rho x_{t-1} + u_t, \tag{13b}$$

where μ is the unconditional mean of the time series $\{y_t\}$, and $\{x_t\}$ is a zero mean AR(1) process. Eliminating x_t one obtains the equivalent representation

$$(y_t - \mu) = \rho(y_{t-1} - \mu) + u_t, \tag{14}$$

which is nonlinear in the parameters μ and ρ. The model can also be written in a form that is linear in the parameters by defining $c = \mu(1-\rho)$, so that

$$y_t = c + \rho y_{t-1} + u_t. \tag{15}$$

Note however that the parameters in (15) are not unrestricted; if $\rho = 1$, then $c = 0$ irrespective of μ. The constant term disappears from (15) under the unit root hypothesis. Whereas we can (and must) have an informative prior distribution on ρ, we would like to specify an uninformative (flat) prior for the constant term (c) or the unconditional mean (μ).

If the priors of c and ρ are assumed independent and uniform, as in DeJong and Whiteman (1989), we could apply the posterior odds analysis for linear regression models discussed in Leamer (1978). This would yield the same simple prescriptions for unit root tests as worked out in section 2. The prior independence assumption for c and ρ is problematic, however. If $\rho = 1$, the interpretation of the constant term changes. For $\rho < 1$, the constant conveys information about the mean of $\{y_t\}$; for $\rho = 1$, it determines the drift of $\{y_t\}$. To exclude a random walk with drift under the null, when a trend is not present under the alternative, the parameter c should shrink to zero if $\rho \to 1$. Such a restriction must be incorporated in the prior.[6] We do not know any conjugate prior that achieves this, which means that we will need numerical methods to compute posterior odds

We prefer to work with the nonlinear specification (14). It has the advantage that μ is an interpretable parameter, so that it seems natural to specify a prior on μ. This prior can always be converted to a prior on the constant c using the transformation $c = \mu(1-\rho)$. The most natural way to proceed would be the formulation of a uniform uninformative prior on μ. As will be shown below this will not work due to the lack of identification of μ under the unit root hypothesis; it cancels from both sides of (14) if $\rho = 1$. With an uniform uninformative prior on μ the posterior probability of the unit root hypothesis will go to unity independent of any data information. Further, a uniform prior implies a discontinuity in the transition from a stationary model to the random walk. These two facts lead us to consider a weakly informative prior on μ that smoothly blends into an uninformative prior as $\rho \to 1$.

In the rest of this section we will represent the AR(1) as in (14). The assumptions of the initial condition y_0, the stochastic process that generates the disturbance $u_1, u_2, \ldots, u_T$, and the *a priori* plausible range of ρ are the same as in section 2. The likelihood function of the parameters (ρ, μ, σ) is of the same functional form as given in eq. (6) of section 2, except that $u'u$ is

[6]The problem remains if a linear trend is included in the stationary case. Then the coefficient of the linear trend must vanish under the unit root in order to exclude a quadratic trend under the random walk. The single unit root restriction always implies two restrictions on a representation that is linear in the parameters.

equal to

$$u'u = [y - \mu(1-\rho)\iota - \rho y_{-1}]'[y - \mu(1-\rho)\iota - \rho y_{-1}], \tag{16}$$

where the $T \times 1$ vectors y and y_{-1} are the same as in section 2 and ι is a $T \times 1$ vector of ones. The parameter σ can still be interpreted as a nuisance parameter. One can integrate the posterior under H_0 and under H_1 with respect to σ by making use of standard results; for details, see section 2. The prior on ρ will also remain unchanged, i.e., uniform on an interval $[a, 1)$.

3.2. *A noninformative prior on* μ

As a preliminary step in the specification of the prior on μ we consider a noninformative uniform prior for μ on a large, though finite, interval:

$$f(\mu) = 1/M, \qquad -M/2 < \mu < M/2. \tag{17}$$

The priors on ρ and σ are the same as in section 2. After integrating the posterior over σ one has

$$\begin{aligned} p(\mu, \rho | Y) &\propto \frac{1-\vartheta}{M(1-a)}(u'u)^{-T/2}, \qquad \rho \in S, \quad \mu \in (-M/2, M/2), \\ &\propto \frac{\vartheta}{M}(u'u)^{-T/2}, \qquad \rho = 1, \quad \mu \in (-M/2, M/2). \end{aligned} \tag{18}$$

The constants of proportionality are the same under the random walk and the stationary AR(1). The top panels ('A') of figs. 4a–4h show the contours of the function $p(\mu, \rho | Y)$ for our real exchange rate data. For $\rho = 1$, the data do not contain information on μ, since the likelihood function does not depend on μ. This creates the 'wall' in the back of the figures. It illustrates that μ does not enter the analysis symmetrically under the null and the alternative, and hence is not a true nuisance parameter like σ. The figures also show the second flaw of a uniform prior on μ. The posterior density is more concentrated along the μ-axis for smaller values of ρ: μ and ρ are not independent. The nonidentification of μ under the unit root hypothesis leads to pathological behaviour of the posterior odds ratio. The result is formulated in the following theorem:

Theorem. Consider a time series $\{y_t\}$ *generated by an AR(1) model with* $\rho \in S \cup \{1\}$ *and normally distributed innovations; and consider priors on* ρ, μ, *and* $\ln \sigma$ *that are uniform and independent with the assumption that the prior on* μ *is defined on the finite interval* $(-M/2, M/2)$, *then the posterior odds diverge to infinity as* $M \to \infty$.

Proof. See appendix.

Intuitively, with $\rho = 1$ the data do not contain information on μ, since the likelihood does not depend on μ. On the other hand, the data will strongly revise the prior distribution of μ for all values of $\rho < 1$, since the likelihood function will be approximately zero for values of μ far from the sample mean. The posterior probability that ρ is in the stationary region is obtained by integrating the likelihood function over the entire space of feasible μ and ρ, using the prior as weight function. The 'averaged' likelihood under the stationary AR(1) will tend to zero, since all points, including the whole large region (see figs. 4a–4h) in which the likelihood is nearly zero, are given equal weight.

3.3. *A weakly informative prior on μ*

Given that μ is not really a nuisance parameter we have to investigate a reasonable class of weakly informative prior distributions. The simplest solution to the problem of specifying a prior on μ would be to specify a finite interval for μ. One could for instance set M in a similar way as we constructed the lower bound a^* in section 2. But a bounded rectangular region for (μ, ρ) does not correspond well with the shape of the likelihood function in figs. 4a–4h. With a large M, integration over μ implies that the likelihood function is integrated over a large region where it has almost no mass (along lines where $\rho \ll 1$). For a small M, however, integration over μ would neglect the mass of the likelihood along lines where ρ is close to unity. As shown above this latter area contributes most to the posterior odds.

The shape of the likelihood function suggests a prior on μ that is specified on a larger region the closer ρ is to unity. Accepting the viewpoint that the posterior should be largely determined by the data this implies that one can not use a prior on μ that is independent of ρ. The prior should state that we can have less knowledge about the unconditional mean of the time series process when it becomes more persistent. One class of conditional priors that incorporates this idea is

$$p(\mu|\rho) = (1-\rho^2)^d / M, \qquad \rho \in S, \quad -\frac{M/2}{(1-\rho^2)^d} < \mu < \frac{M/2}{(1-\rho^2)^d}. \tag{19}$$

For $d = 0$, we have our previous prior defined on a bounded rectangular region. For all d the conditional prior on μ is still uniform, though the range of μ varies with ρ. To implement this prior one has to specify values for d and M. The value of M can be determined empirically or as a function of the innovation variance σ (for example $M = 3\sigma$). A reasonable choice for d is less obvious, and we will not pursue this further here.[7]

Instead we will utilize the unconditional distribution of y_t when $\{y_t\}$ is a stationary time series to specify a weakly informative prior on μ. If the time series $\{y_t\}$ is stationary, the unconditional mean and variance of the series are given by $\mathrm{E}(y) = \mu$ and $\mathrm{var}(y) = \sigma^2/(1-\rho^2)$. Assuming normality, the unconditional distribution of y_0 will be normal with parameters μ and $\sigma^2/(1-\rho^2)$. Hence, conditional on the initial condition y_0 and the parameters ρ and σ, a reasonable prior on μ is given by

$$p(\mu \mid \rho, \sigma, y_0) \sim \mathrm{Normal}(y_0, \sigma^2/(1-\rho^2)), \qquad \rho \in S. \tag{20}$$

This prior is weakly informative. Contrary to a uniform prior on the entire real line, the prior in (20) is centered around y_0, and has a variance that is determined by the other parameters that describe the time series process. The prior is stronger the smaller the value of ρ. The prior approaches an uninformative prior if $\rho \to 1$, since the variance of μ then goes to infinity. This reflects the fact that a random walk does not have an unconditional mean. The prior also fits well with the shape of the likelihood function (see figs. 4a–4h). The prior becomes flatter when $\rho \to 1$, as does the likelihood. Technically this normal prior has the same effect as setting $d = \frac{1}{2}$ in (19).

The prior for μ at $\rho = 1$ can be any distribution, as it does not affect the posterior odds. The priors on ρ and σ are the same as before. The derivation of the odds proceeds analogously to the computations in section 2. The details can be found in the appendix. Here we state the result:

$$K_1 = \frac{\vartheta}{1-\vartheta} \cdot \frac{(\Delta y' \Delta y)^{-T/2}}{\dfrac{1}{1-a}\displaystyle\int_a^1 h(\rho) S^2(\rho)^{-T/2}\,\mathrm{d}\rho}. \tag{21}$$

where

$$h(\rho) = \left(1 + T\frac{1-\rho}{1+\rho}\right)^{-1/2},$$

$$\hat{\mu}(\rho) = \frac{\bar{y} - \rho\bar{y}_{-1} + y_0(1+\rho)/T}{1-\rho+(1+\rho)/T},$$

$$S^2(\rho) = (y - \iota\hat{\mu}(\rho)(1-\rho) - \rho y_{-1})'(y - \iota\hat{\mu}(\rho)(1-\rho) - \rho y_{-1}) + (1-\rho^2)(y_0 - \mu(\hat{\rho}))^2.$$

[7]A similar prior, with $d = 0.5$, is suggested by Zellner (1971, ch. 7), who derives it from the determinant of the information matrix from the unconstrained linear representation (15).

The expression for $\hat{\mu}(\rho)$ is a weighted average of the usual OLS estimator of μ (as a function of ρ) and the prior mean y_0. $S^2(\rho)$ denotes the residual sum of squares as a function of ρ. Since $S^2(1)=\Delta y'\Delta y$, the integrand $h(\rho)S^2(\rho)^{-T/2}$ has a continuous transition to $(\Delta y'\Delta y)^{-T/2}$, which appears in the numerator of the odds. The posterior odds are no longer an exact function of the Dickey–Fuller $\hat{\tau}_{\mu}$ statistic; differences are due to the treatment of the first observation and the role of the normal prior.

4. Unit roots in real exchange rates

We have selected eight different real exchange rates. Most of them are exchange rates of the *US dollar* against the currency of another developed country: West Germany, Japan, United Kingdom, France, Canada, and the Netherlands. For comparison we also included two exchange rates against the German mark of countries that are a member of the European Monetary System (EMS): France and the Netherlands. The raw data are monthly time series of nominal exchange rates and consumer price indices obtained from the IFS databank (series *ae* and *64*) over the sample 73:01 to 88:07, except for United Kingdom and Canada where the final value is 88:06. All variables are transformed to logarithms. Real exchange rates were constructed as $y=e-p+p^*$, where e is the log nominal exchange rate expressed as the domestic price of one unit of foreign currency, and p and p^* are the logarithms of the consumer price index of the domestic and foreign country respectively. Graphs of all series are shown in fig. 3.

The notion of Purchasing Power Parity (PPP) is fundamental in the linkage of international price levels. It maintains that one dollar can buy approximately the same amount of goods all over the world. The ratio of the US price level to the foreign price level, converted to dollars by the nominal exchange rate, should be equal to one in equilibrium. As prices are measured as indices with respect to some base year, the observed ratio must be equal to some constant.

As a long-run equilibrium condition PPP does not pose restrictions on the time series process of real exchange rates other than a constant unconditional mean. The existence of a unit root, and in particular a random walk model, would contradict this. In investigating the PPP hypothesis we compare the random walk model, which has often been found as a good description of the data, to the alternative of a stationary AR(1). In a preliminary screening of the data we found no evidence of a linear trend except in certain subsamples that stopped in 1985, the peak of the five-year real appreciation of the dollar against most currencies (see the plotted data series). Also higher-order dynamics do not appear to be present. Some appearance of twelfth-order autocorrelation points to a seasonal pattern in some of the price indices. This seasonal effect is very small, though, compared to the

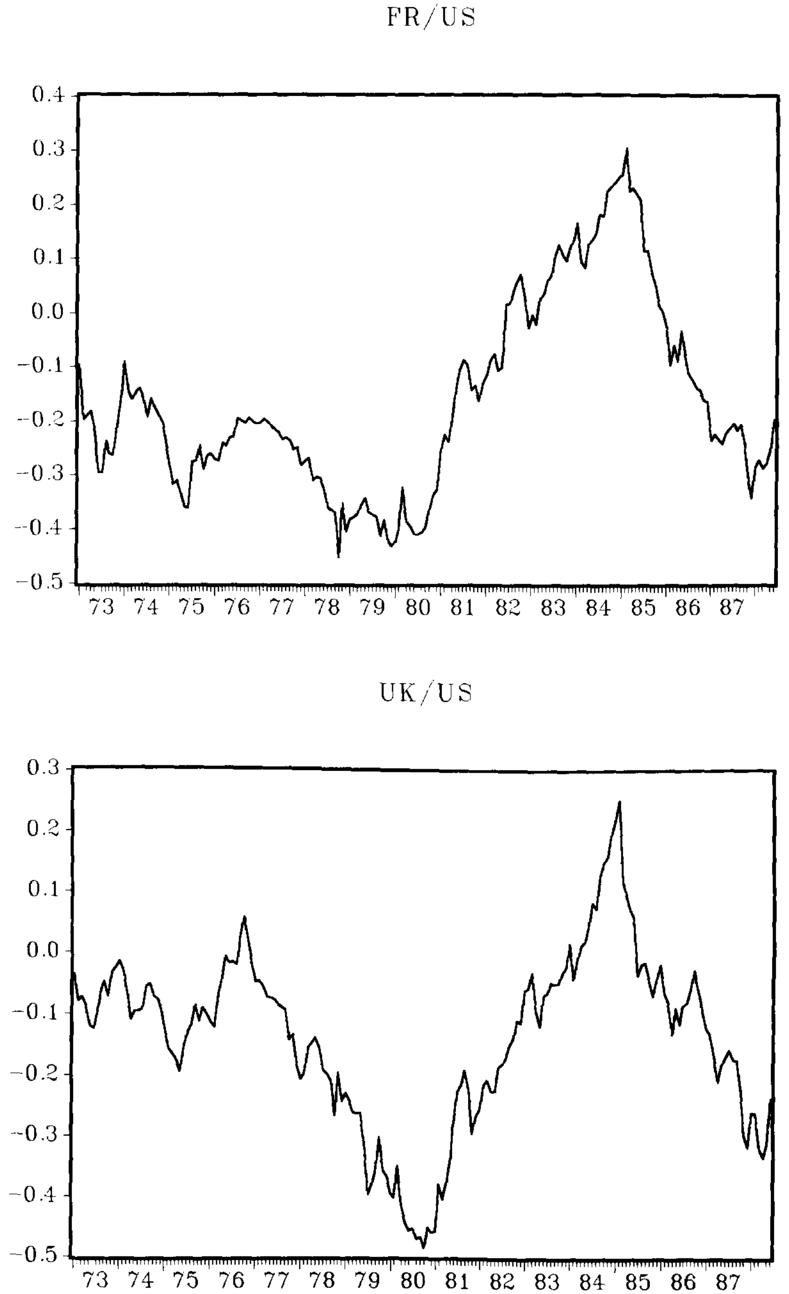
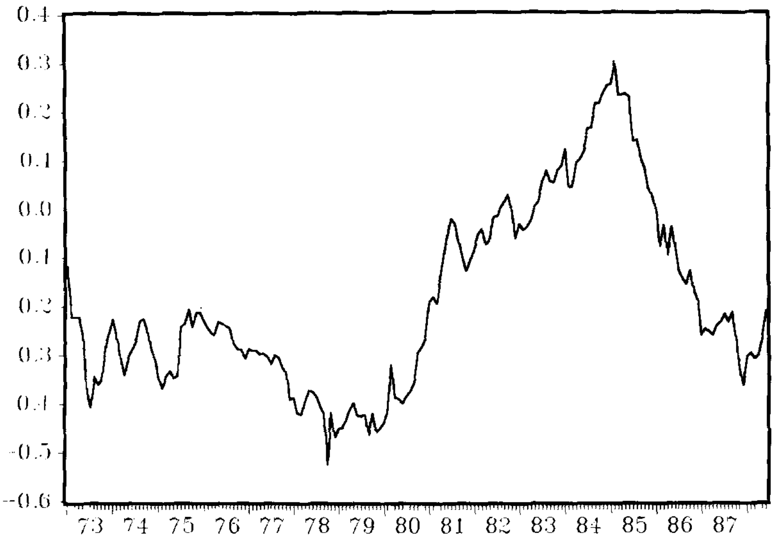


Fig. 3. Log of real exchange rates (72:12–88:6).

WG/US



NL/US

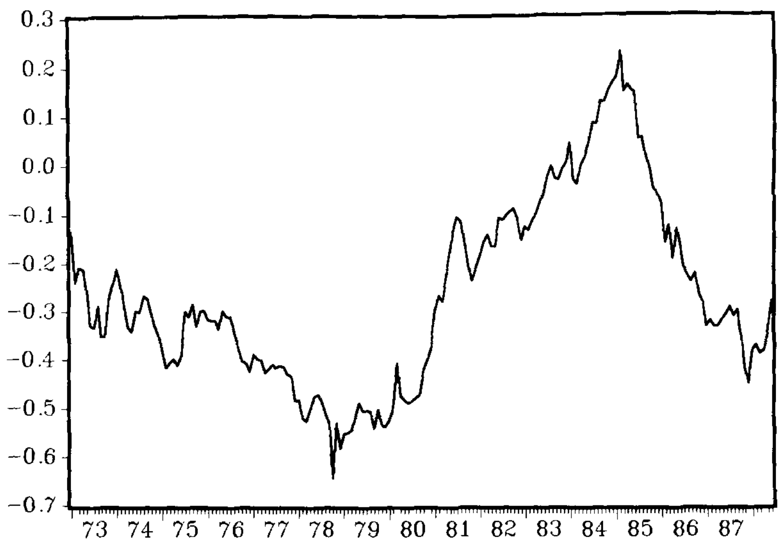
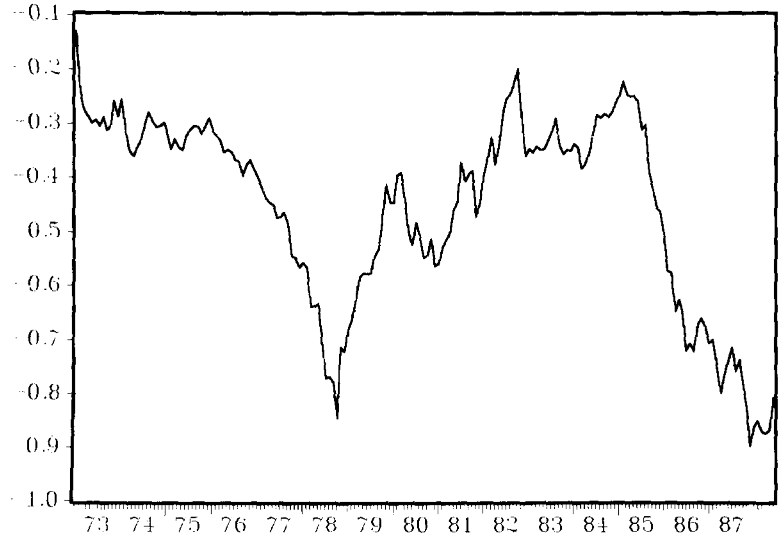


Fig. 3 (continued)

JP/US



FR/WG

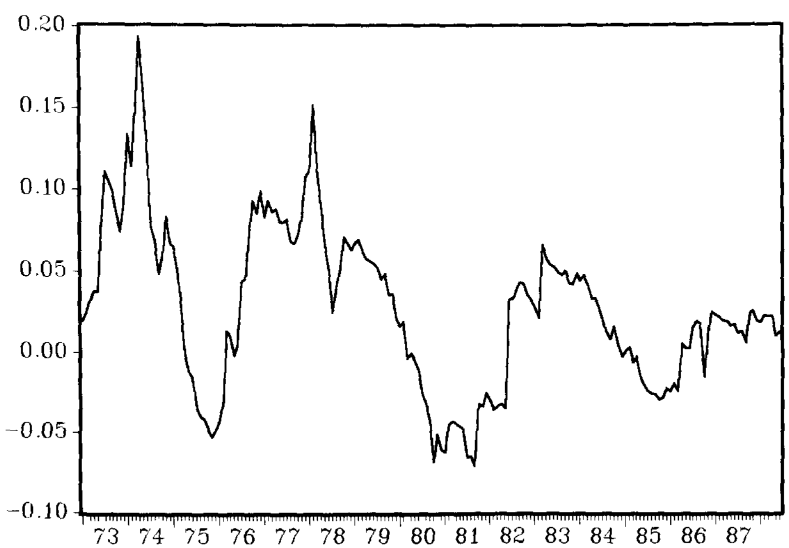
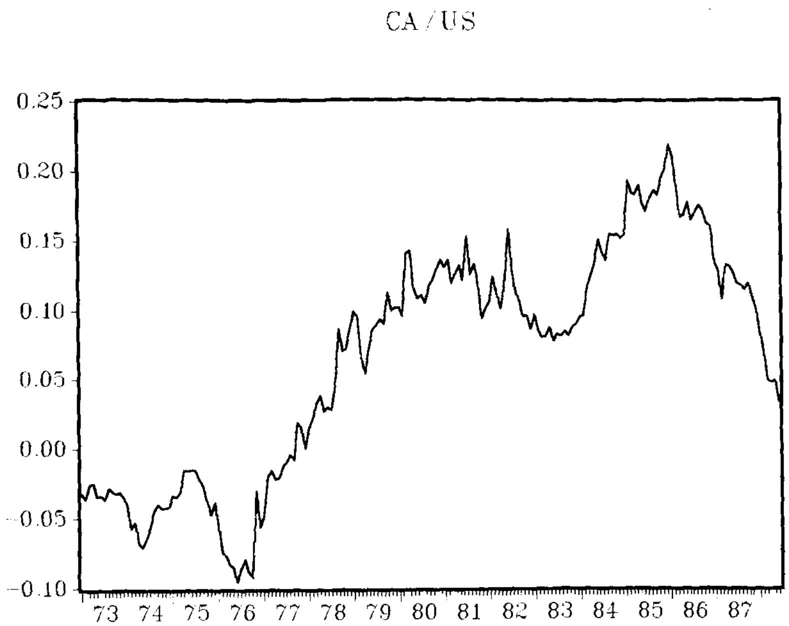


Fig. 3 (continued)



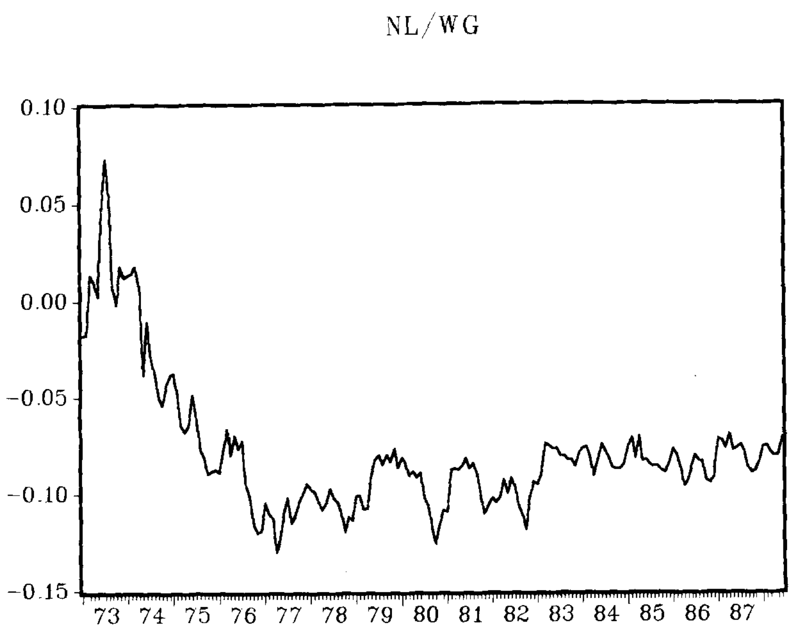


Fig. 3 (continued)

Table 1

Unit root tests of real exchange rates.[a]

	Dickey–Fuller $\hat{\tau}_\mu$	Bhargava R_1	Posterior odds with constant	Probability $\rho = 1$
FR/US	−1.34	0.036	1.16	0.54
WG/US	−1.33	0.036	1.16	0.54
JP/US	−1.25	0.037	2.81	0.74
CA/US	−1.38	0.027	1.55	0.61
UK/US	−1.62	0.054	1.05	0.51
NL/US	−1.41	0.036	1.21	0.55
FR/WG	−2.17	0.101	0.39	0.28
NL/WG	−2.40	0.089	0.75	0.43

[a]Country codes: US = United States, UK = United Kingdom, WG = West Germany, FR = France, NL = Netherlands, JP = Japan, CA = Canada.

variation in exchange rates. These results induced us to analyze unit roots within the class of AR(1) models.

Test results on the unit root hypothesis are reported in table 1. The first two columns contain classical unit root tests: the Dickey–Fuller test $\hat{\tau}_\mu$ [see Fuller (1976)] and Bhargava's (1986) R_1 test. The null hypothesis of a unit root can not be rejected for any of the currencies at the 5% or 10% level with these tests. These results confirm the random walk results in the literature. The surprising result is that one cannot even reject unit root for the currencies that are in the EMS (NL/WG and FR/WG). Although the series of the German mark against the US dollar and the French franc against the dollar are very similar, the cross-rate obtained by subtracting the two series is still not distinguishable from a random walk.

One caveat with these results is the possible presence of ARCH and nonnormality in the residuals. Diagnostic tests[8] revealed that for two series, NL/US and NL/WG, the residuals from the AR(1) regression show significant ARCH type heteroskedasticity. Normality was rejected for all series except FR/US and NL/US. We therefore computed unit root test statistics using the Perron (1988) correction of the Dickey–Fuller $\hat{\tau}_\mu$-statistic. The corrected t-values of ρ did not differ much from those in the table and did not affect conclusions. A detailed analysis of the robustness of our test results with respect to these misspecifications is outside the scope of this paper. We conjecture [with Sims (1988)] that they will not affect the results greatly.

Column 3 presents posterior odds ratios; column 4 the corresponding posterior probability that the series is generated by a random walk.[9] They have been computed as described in section 3 for the model with a constant

[8]The diagnostics are not reported, but are available on request.

[9]The odds differ from the results in a preliminary version of the paper due to a numerical inaccuracy in the computation of the cumulative marginal posterior of ρ. The integration routine

term using the normal prior. The most marked difference with the classical tests occurs for the French franc/German mark (FR/WG) real exchange rates. Though the random walk is not rejected at the 5% or 10% level by the two classical tests, the posterior odds are well below one. Only one series comes out as a random walk: JP/US. For the other currencies (FR/US, WG/US, CA/US, UK/US, NL/US, NL/WG) the posterior probabilities are close to one half, implying that the random walk and a stationary AR(1) are about equally likely. The only two posterior odds ratios below one are obtained for NL/WG and FR/WG, the two EMS exchange rates in our dataset.

The effect of the normal prior on μ can be seen by comparing panels A and B in figs. 4. Panel A (the top figure) in these figures shows the bivariate posterior of μ and ρ with a flat prior; panel B (at the bottom) shows the posterior with the normal prior, conditional on $\rho \in S$. The relatively large amount of probability mass close to the unit root in panels A reflects the nonintegrability which causes the odds to become infinite, as explained in section 3. The posteriors in panel B are much more concentrated due to the informative prior. The nonlinear dependence between μ and ρ is present with both priors; the conditional distributions of μ given ρ have wider tails the closer ρ moves to the unit root.

The marginal posterior of ρ obtains after integrating the bivariate posterior over μ. The probability density of the bivariate posterior is high and spread out close to $\rho = 1$. Hence integration over μ yields a relatively high univariate density for large values of ρ. This is visualized in fig. 5, which shows the marginal posteriors of ρ for all eight exchange rate series. The vertical line labeled ML in these figures indicates the location of the mode of ρ in the bivariate posterior. Since the bivariate posterior has the functional form of the exact likelihood of an AR(1) model, this mode is equal to the exact ML estimate of ρ. The mode of the marginal density is always to the right of the ML estimate. This illustrates the difference between concentration and marginalization with the skewed distributions that one encounters for the real exchange rate time series. A complete set of the parameter estimates is reported in table 2.

From the analysis in section 3 we know that for $\vartheta = 0.5$ the posterior is continuous in ρ. So the posterior odds are the values of the posterior on the $\rho = 1$ axis in the figures divided by the average area under the posterior density (with a cutoff point when 99% of the area is covered; see section 2 for the construction of a^*). In the case of the French franc/German mark (FR/WG in fig. 5) the posterior is relatively small at $\rho = 1$, compared to the other currencies. Almost the mass of the posterior is to the left of $\rho = 1$. Hence the posterior odds in table 1 are very small. This contrasts enormously

used too few function evaluations. This also affects the graphs of the univariate marginal densities which were plotted using these evaluations. All programs are written in GAUSS 2.0.

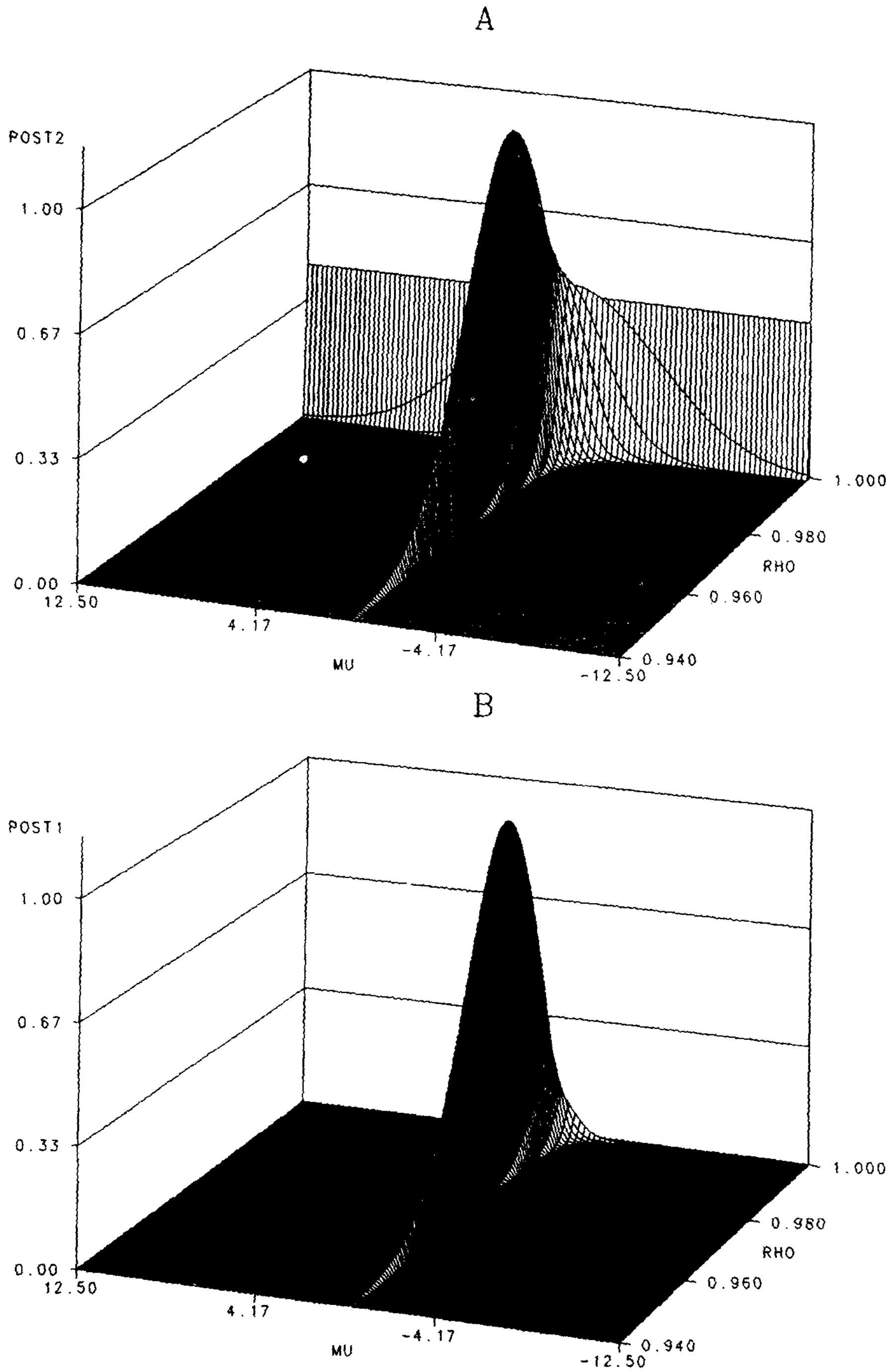
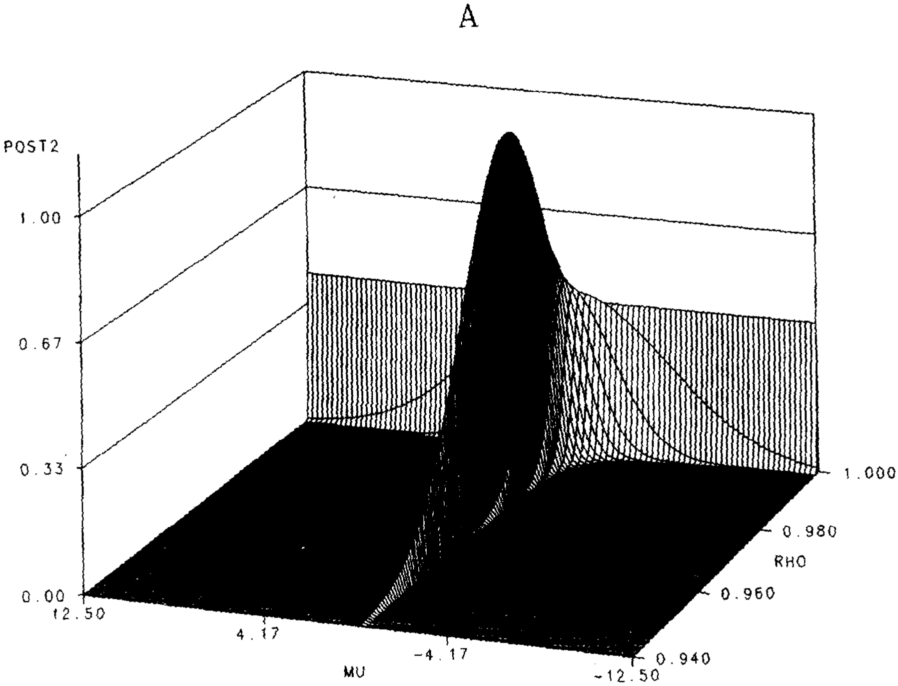


Fig. 4a. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for FR/US.



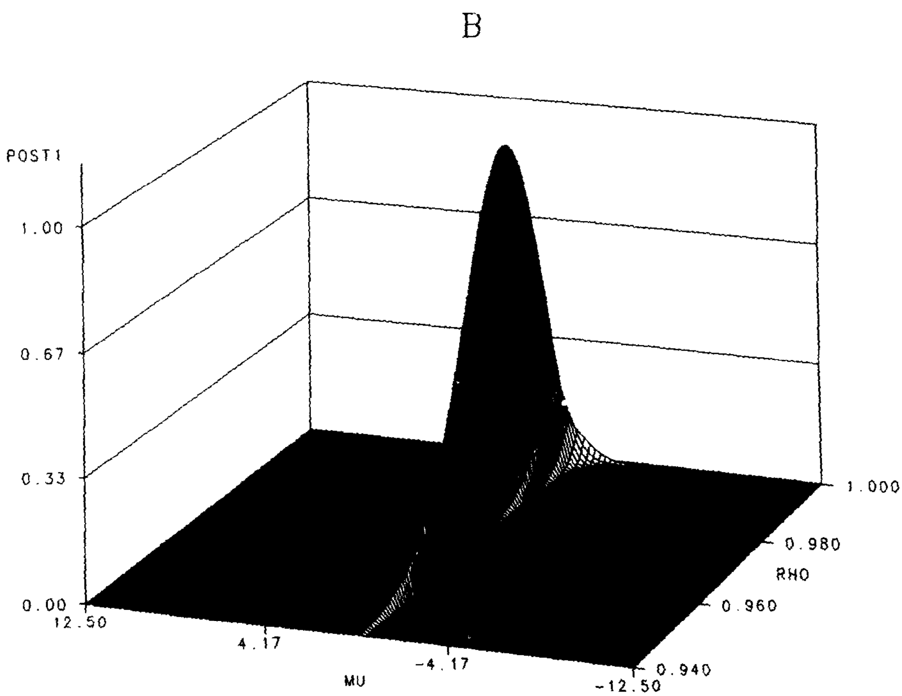


Fig. 4b. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for WG/US.

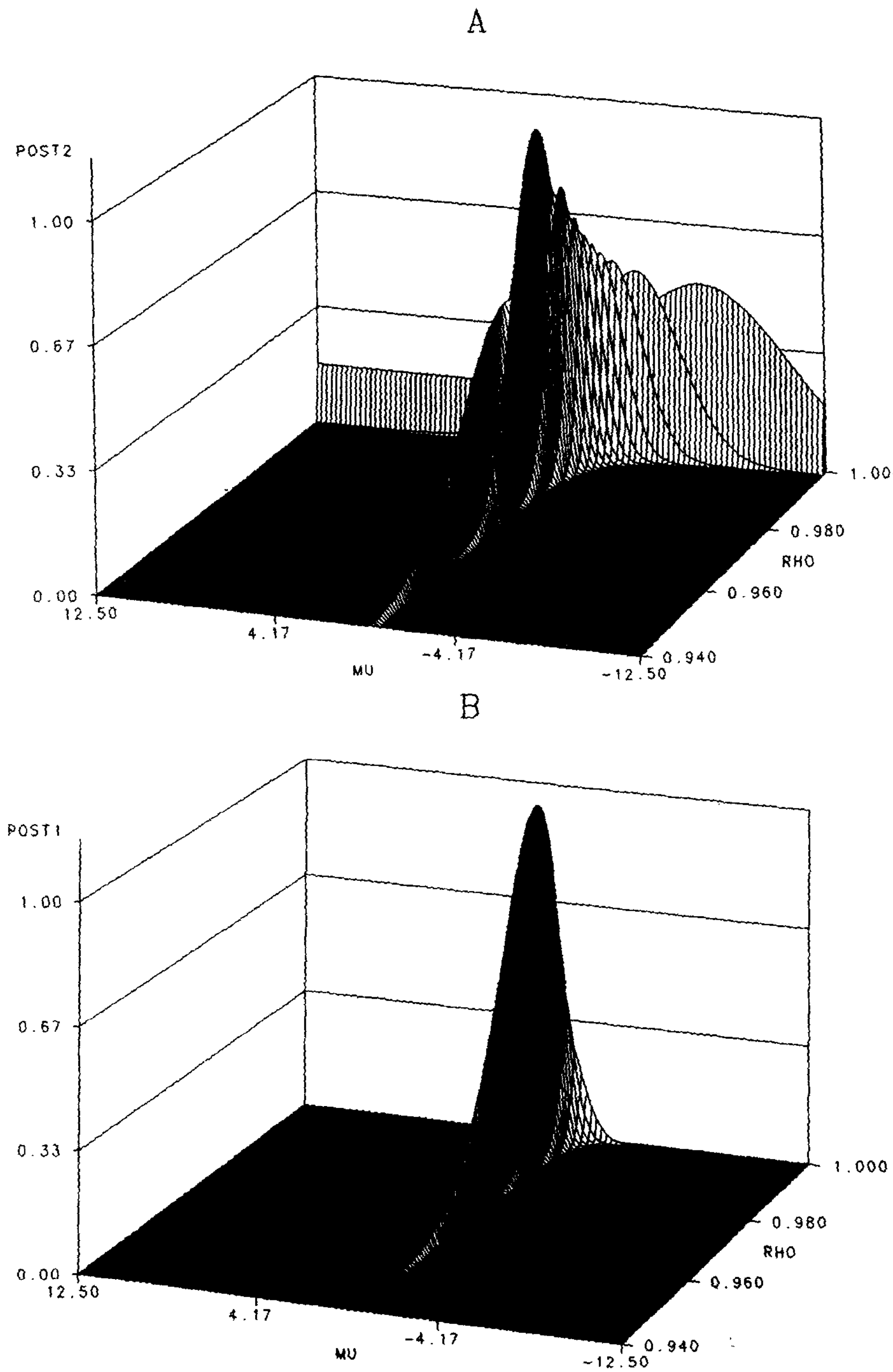
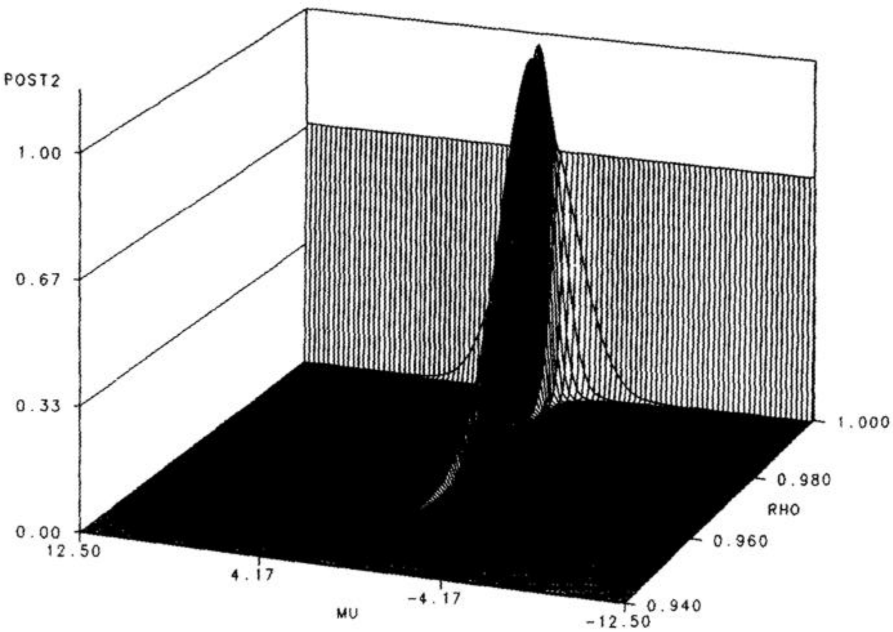


Fig. 4c. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for JP/US.

A



B

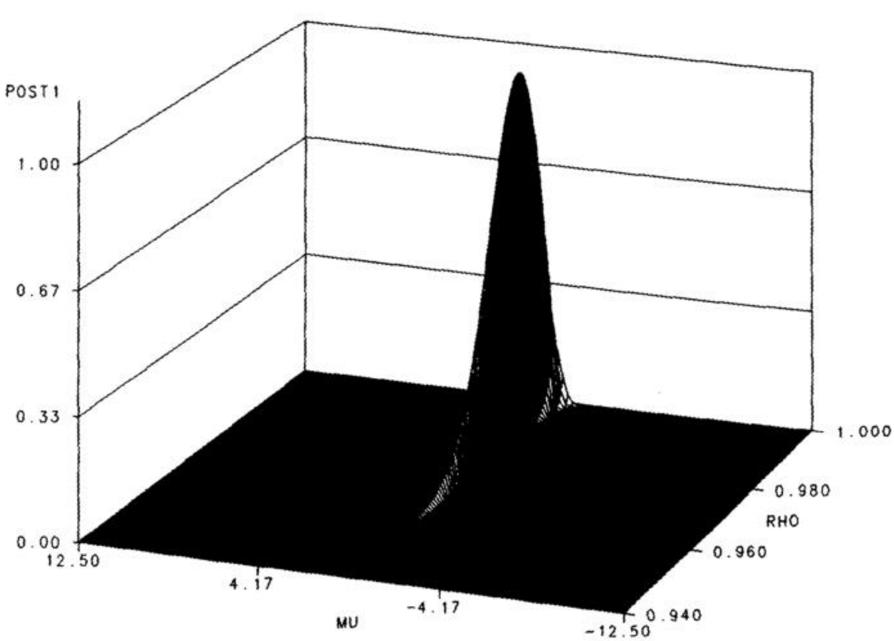


Fig. 4d. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for CA/US.

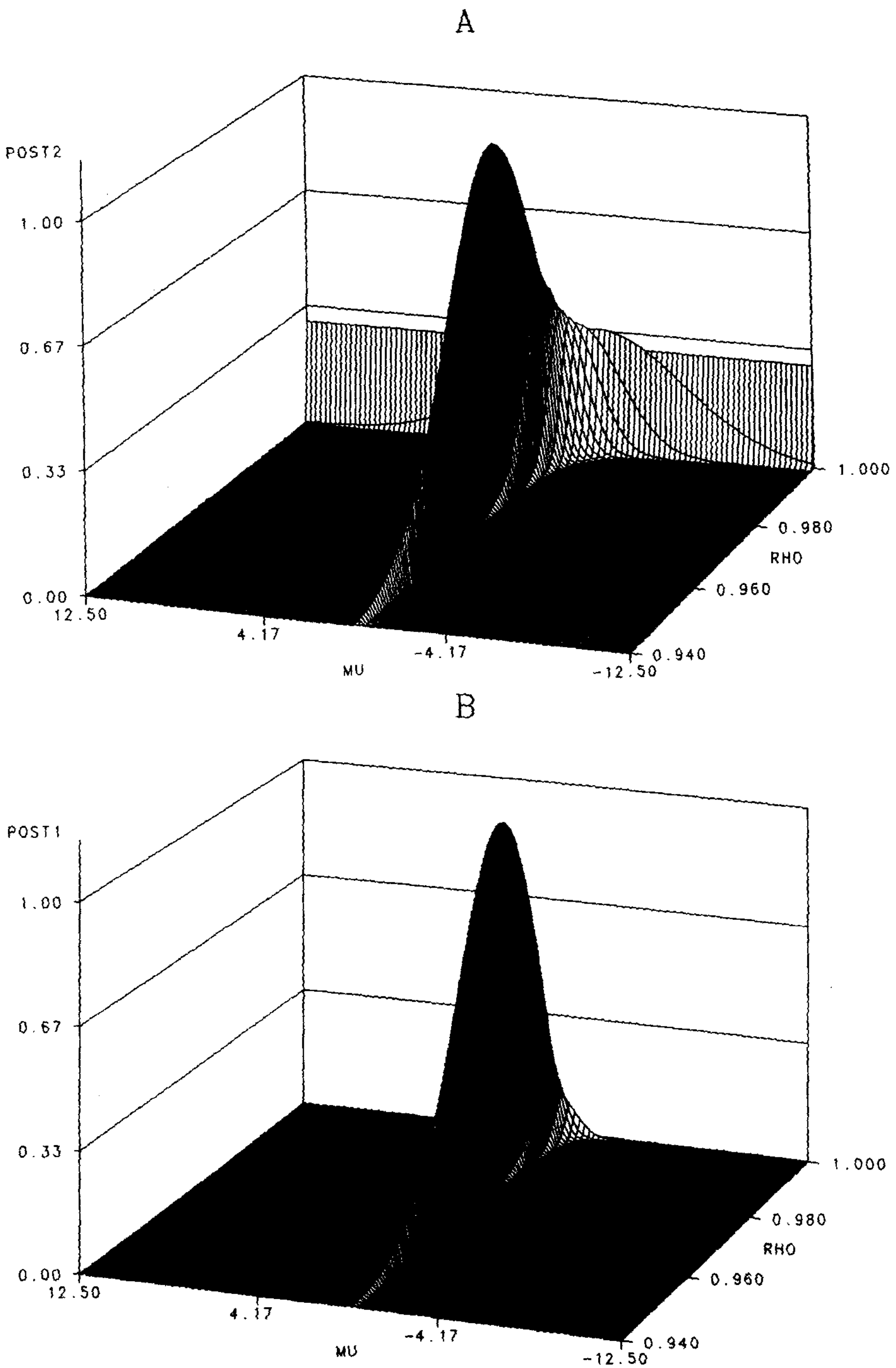
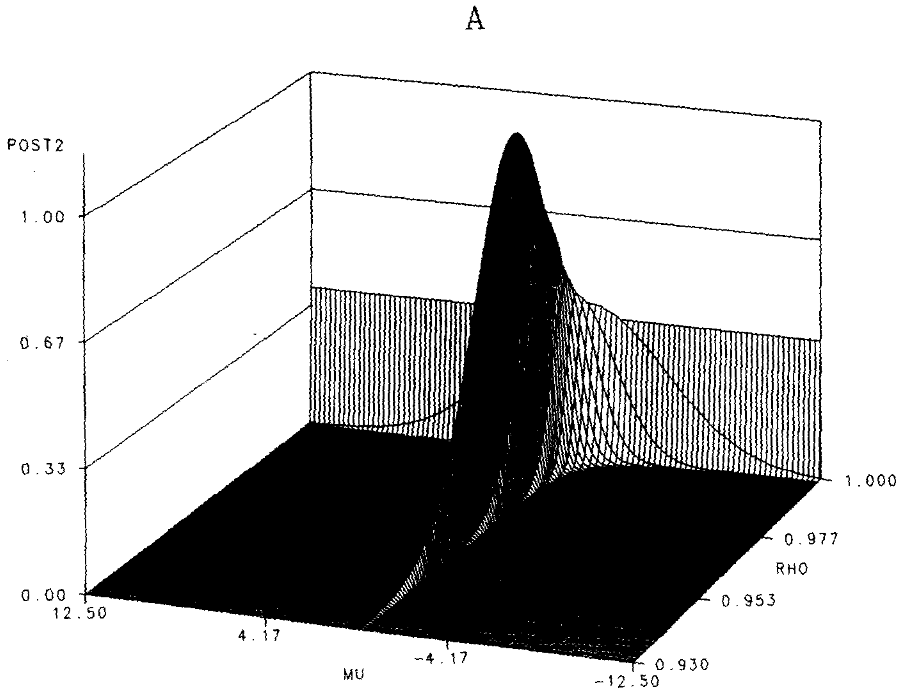


Fig. 4e. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for UK/US.



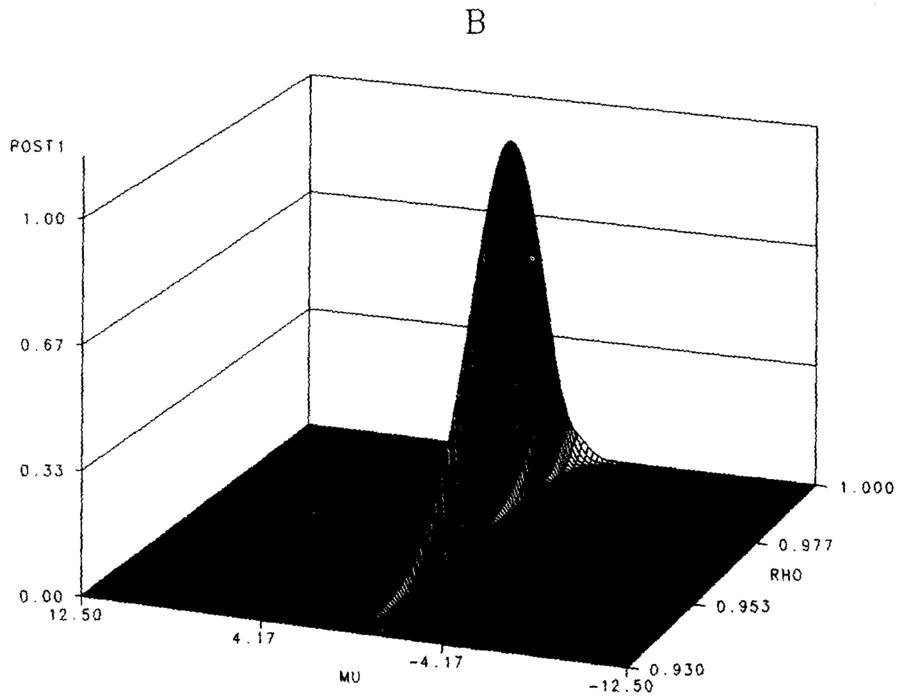


Fig. 4f. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for NL/US.

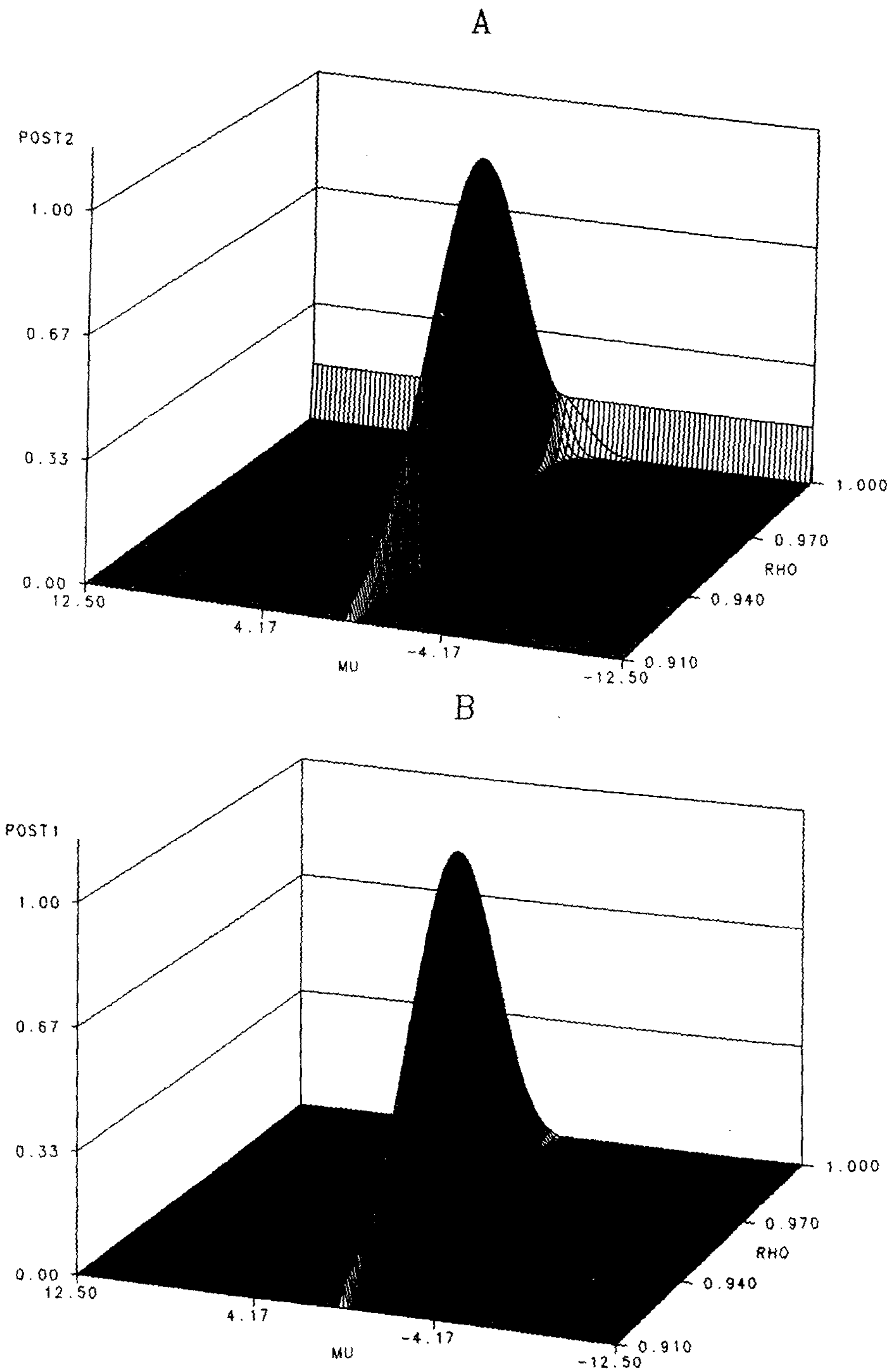


Fig. 4g. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for FR/WG.

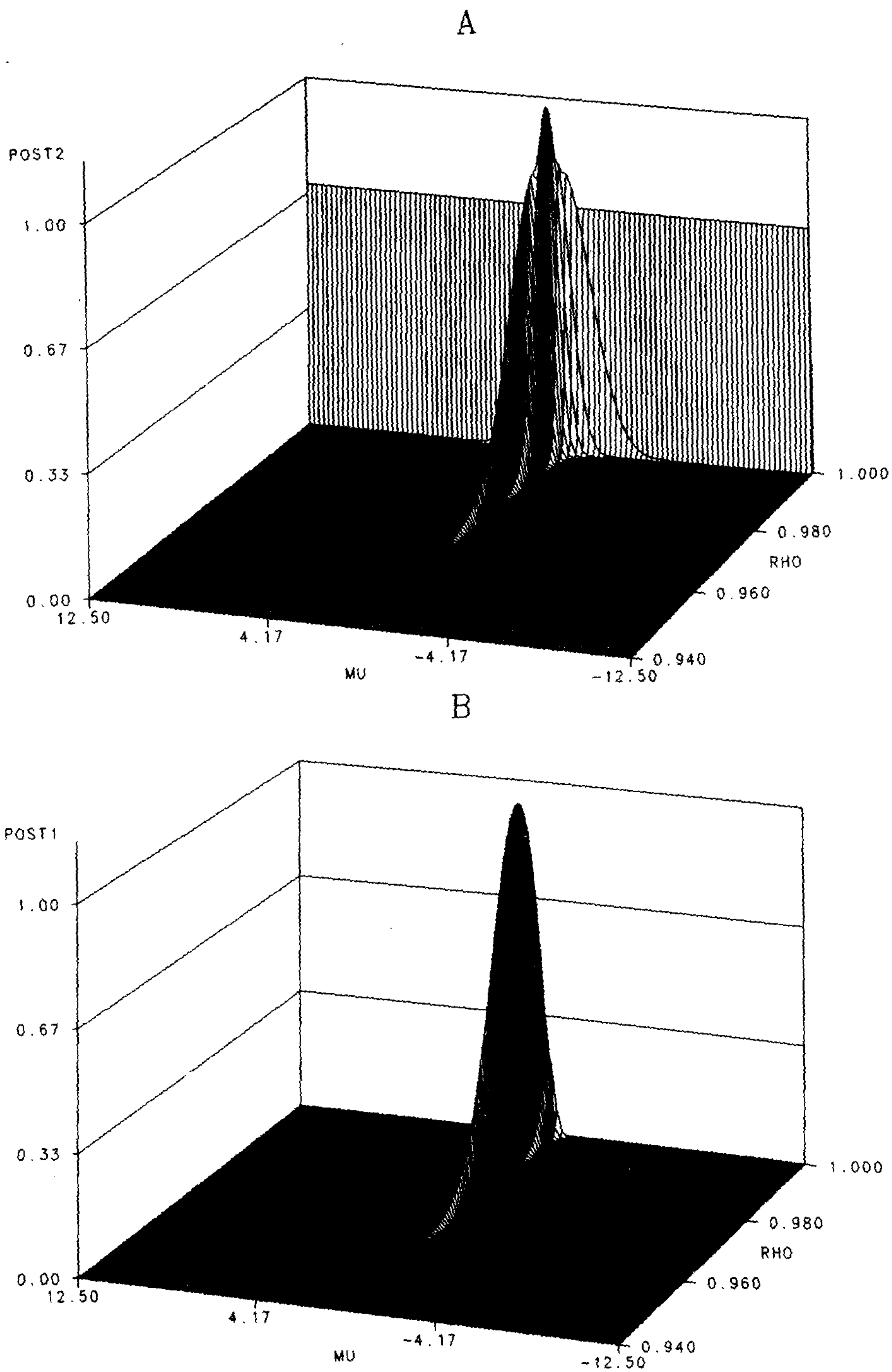


Fig. 4h. Bivariate posterior of (μ, ρ) with uniform prior (panel A) and normal prior (panel B) for NL/WG.

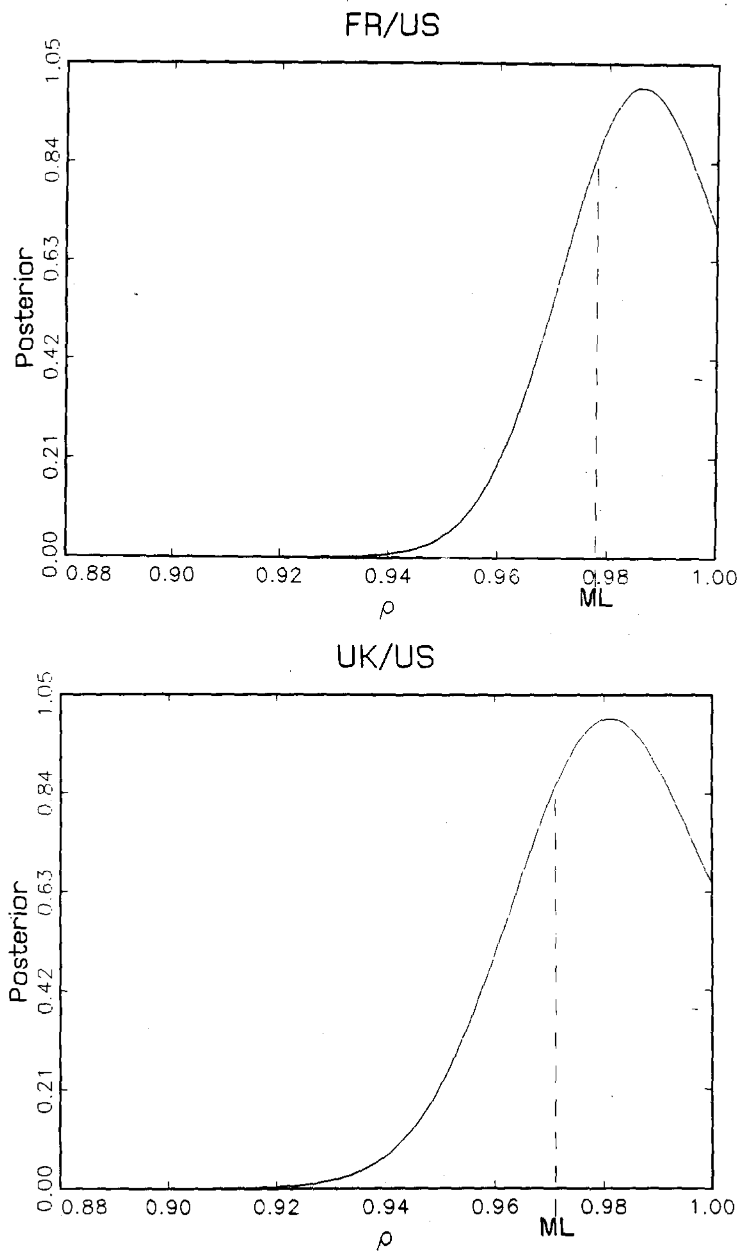


Fig. 5. Univariate marginal posterior of ρ.

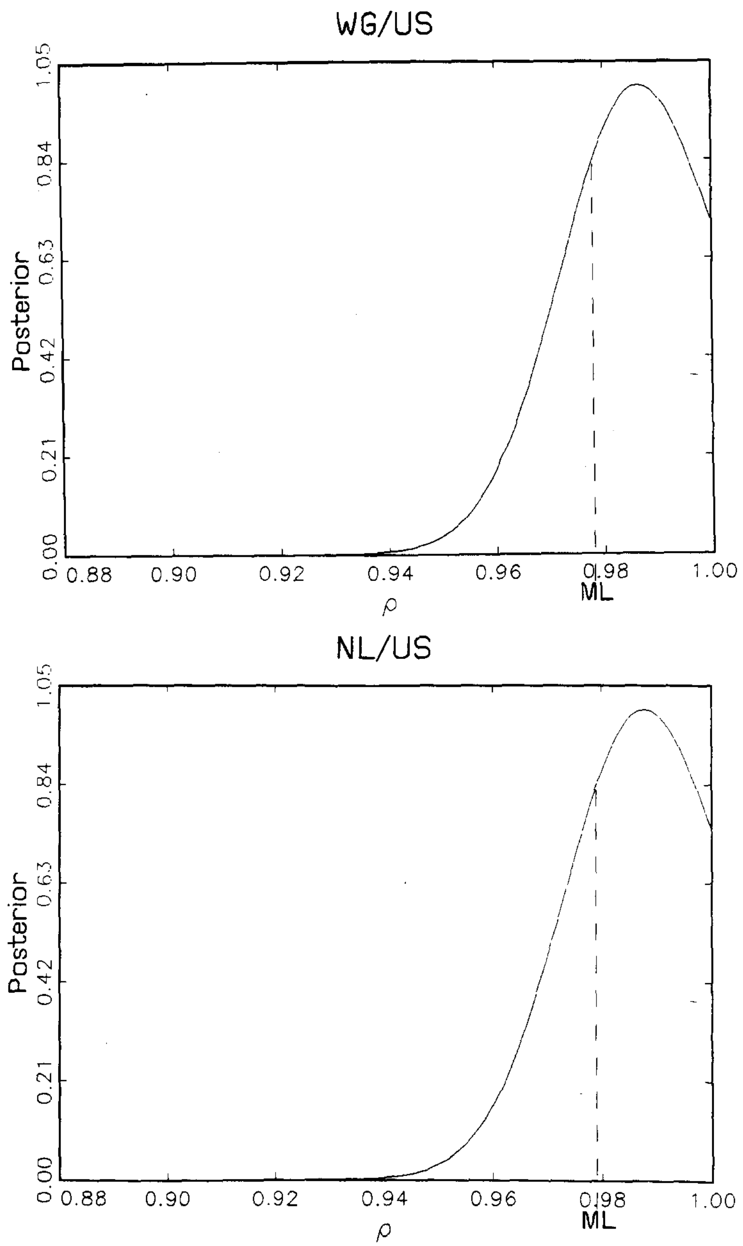


Fig. 5 (continued)

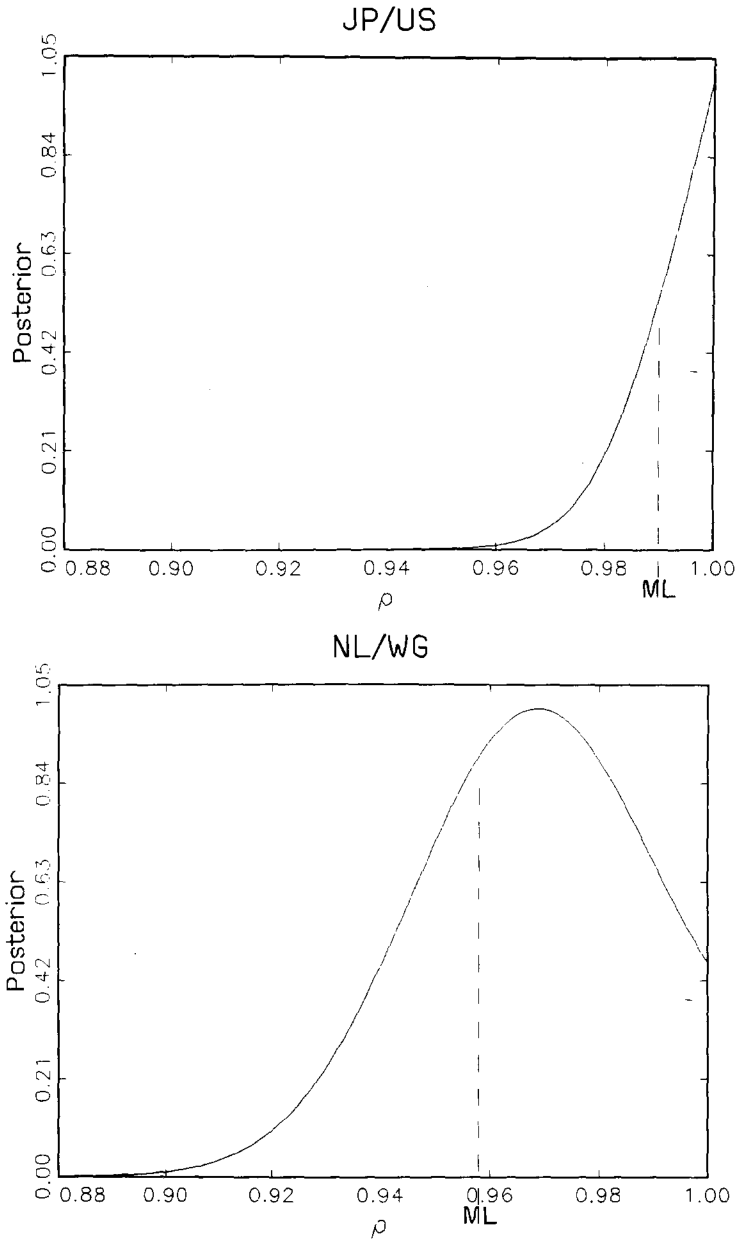


Fig. 5 (continued)

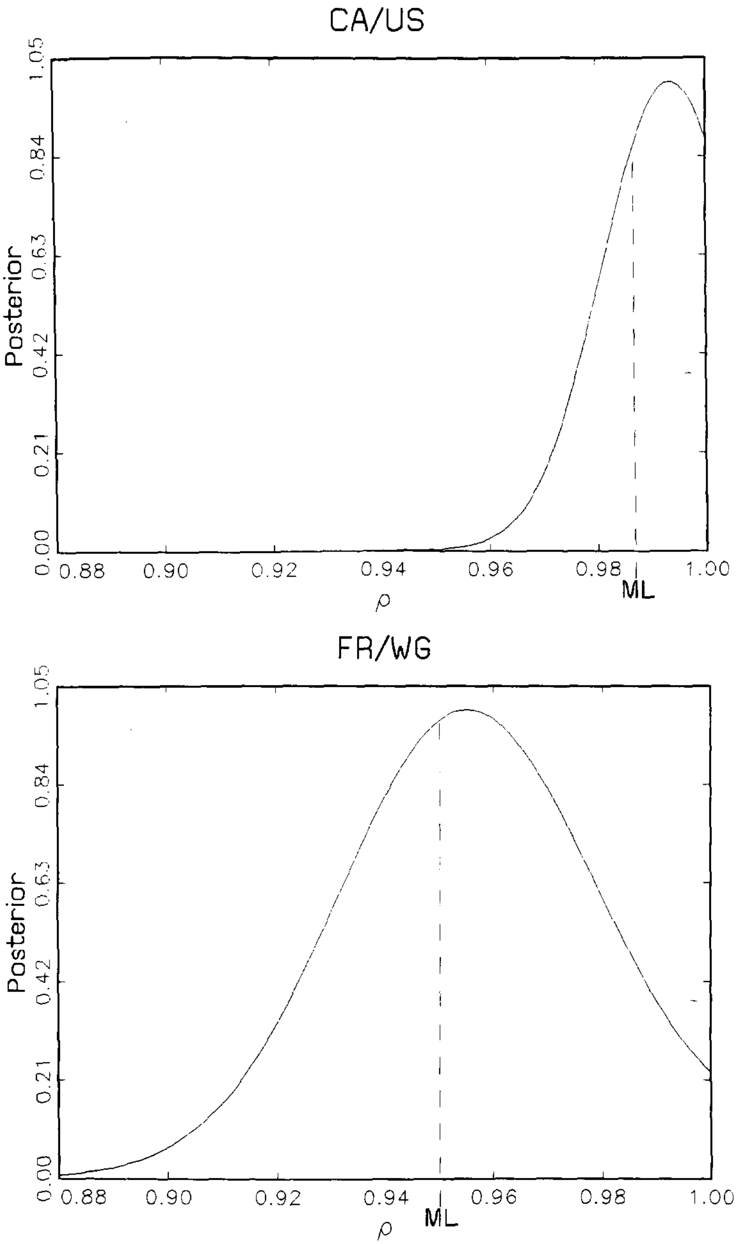


Fig. 5 (continued)

Table 2
Parameter estimates of AR(1) model.[a]

	$E(\rho)$	$\hat{\rho}_{\mathrm{ML}}$	$\hat{\rho}_{\mathrm{OLS}}$	$1-a^*$
FR/US	0.982 (0.012)	0.978 (0.013)	0.981 (0.014)	0.050
WG/US	0.982 (0.011)	0.978 (0.012)	0.982 (0.014)	0.049
JP/US	0.990 (0.008)	0.990 (0.010)	0.982 (0.014)	0.035
CA/US	0.987 (0.009)	0.985 (0.010)	0.983 (0.012)	0.038
UK/US	0.977 (0.014)	0.971 (0.016)	0.973 (0.017)	0.062
NL/US	0.983 (0.011)	0.979 (0.012)	0.981 (0.014)	0.047
FR/WG	0.954 (0.022)	0.950 (0.022)	0.950 (0.023)	0.100
NL/WG	0.965 (0.020)	0.958 (0.020)	0.948 (0.021)	0.085

[a] $E(\rho)$ denotes the posterior expectation of ρ conditional on $\rho < 1$; $\hat{\rho}_{\mathrm{ML}}$ is the exact maximum likelihood estimate of ρ; $\hat{\rho}_{\mathrm{OLS}}$ is the least squares estimate of ρ. The length of the interval of the posterior used in the computation of the posterior odds is $(1-a^*)$. Standard deviations are in parentheses.

with the posterior of the yen/dollar real exchange rate (JP/US in fig. 5), which still hasn't reached its mode when it hits the $\rho = 1$ axis. The maximum value of the posterior in the interval $[0,1]$ is at $\rho = 1$. Hence the averaged posterior over any subinterval $[a,1]$ will always be smaller than the posterior in $\rho = 1$, and thus the odds will favour the random walk. The other figures are intermediate cases. The posterior mode is attained for $\rho < 1$. It then depends on how far away from unity the mode is located, and on how steep the posterior falls toward the $\rho = 1$ axis, to determine the posterior odds. The flatter the posterior, the more likely the odds will favour the random walk.

The sensitivity of the odds with respect to the length of the prior interval of ρ is shown in fig. 6. This set of figures show the posterior odds as a function of the lower bound a. The empirical lower bound is labelled a^* in these figures. They are just a transformation of the marginal posterior densities shown in fig. 5. For all series, except JP/US there exists a lower bound for which the posterior odds favour stationarity. For JP/US the odds are always larger than one, due to the fact that the posterior attains its mode at the boundary $\rho = 1$. The figures show that one must have a tight prior on ρ in the stationary region in order to reject the random walk. Someone who has a prior that is spread out over the full range $[0,1)$ would be very embarrassed

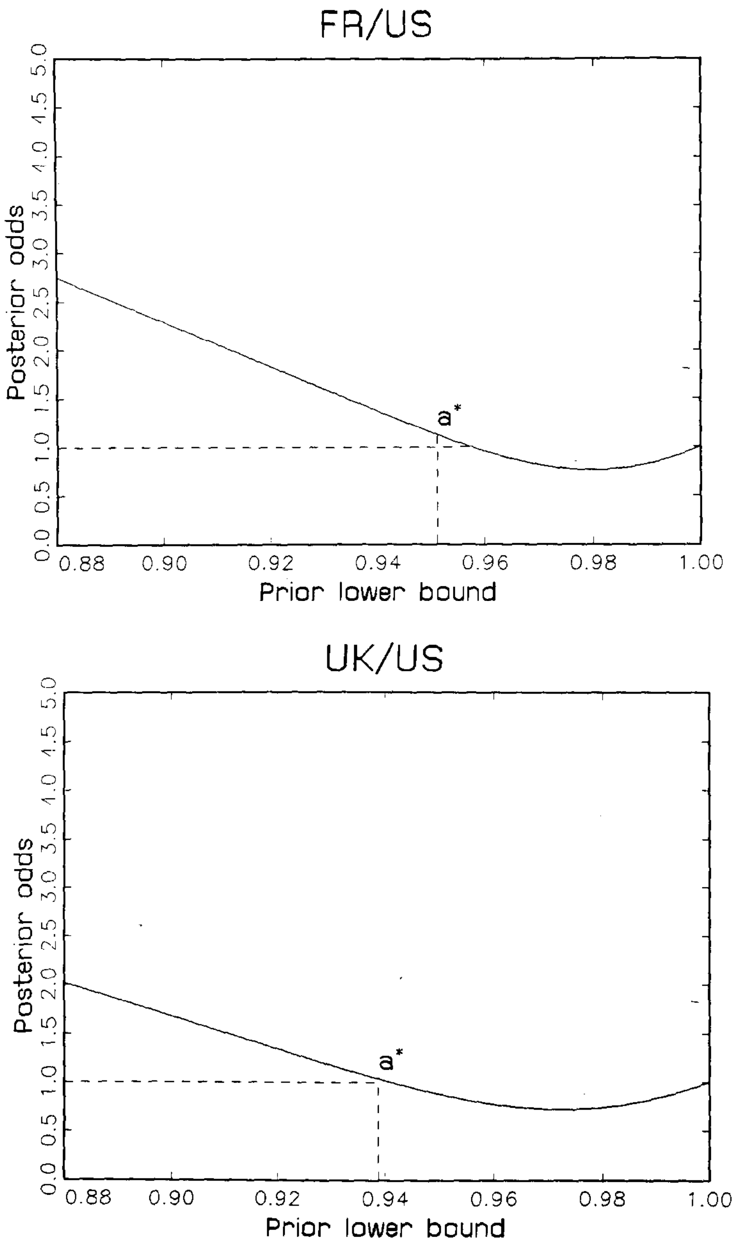
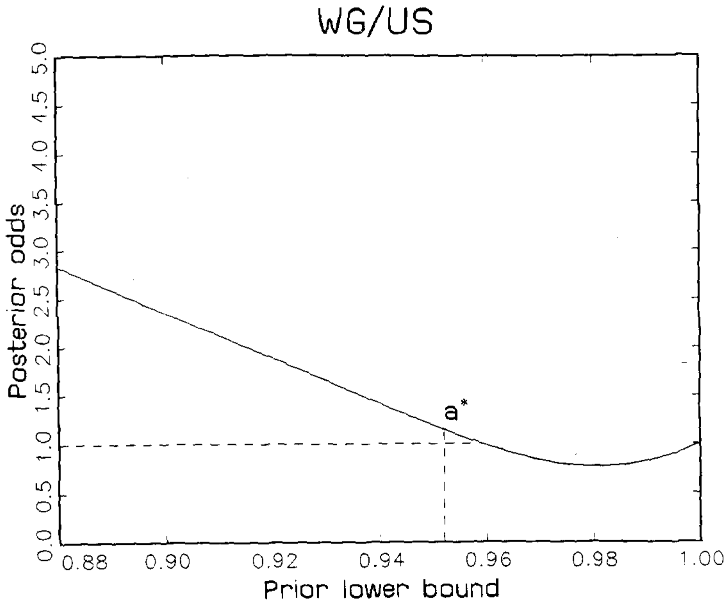


Fig. 6. Sensitivity of posterior odds with respect to prior lower bound.



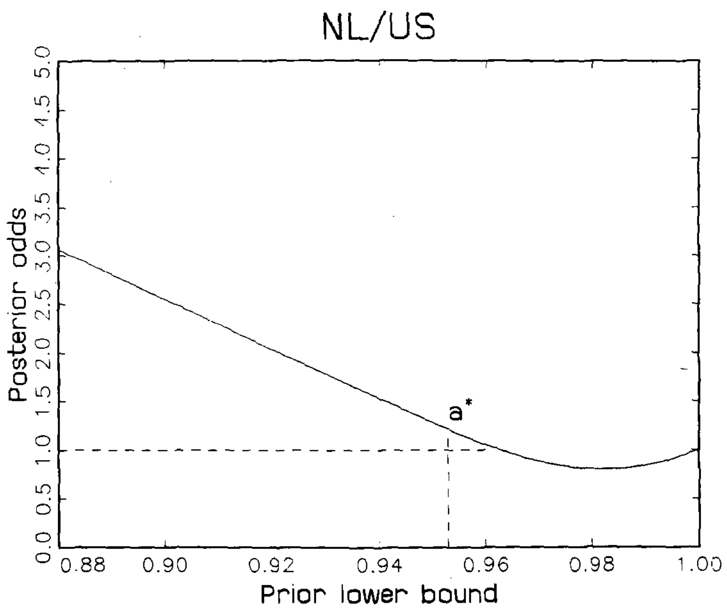


Fig. 6 (continued)

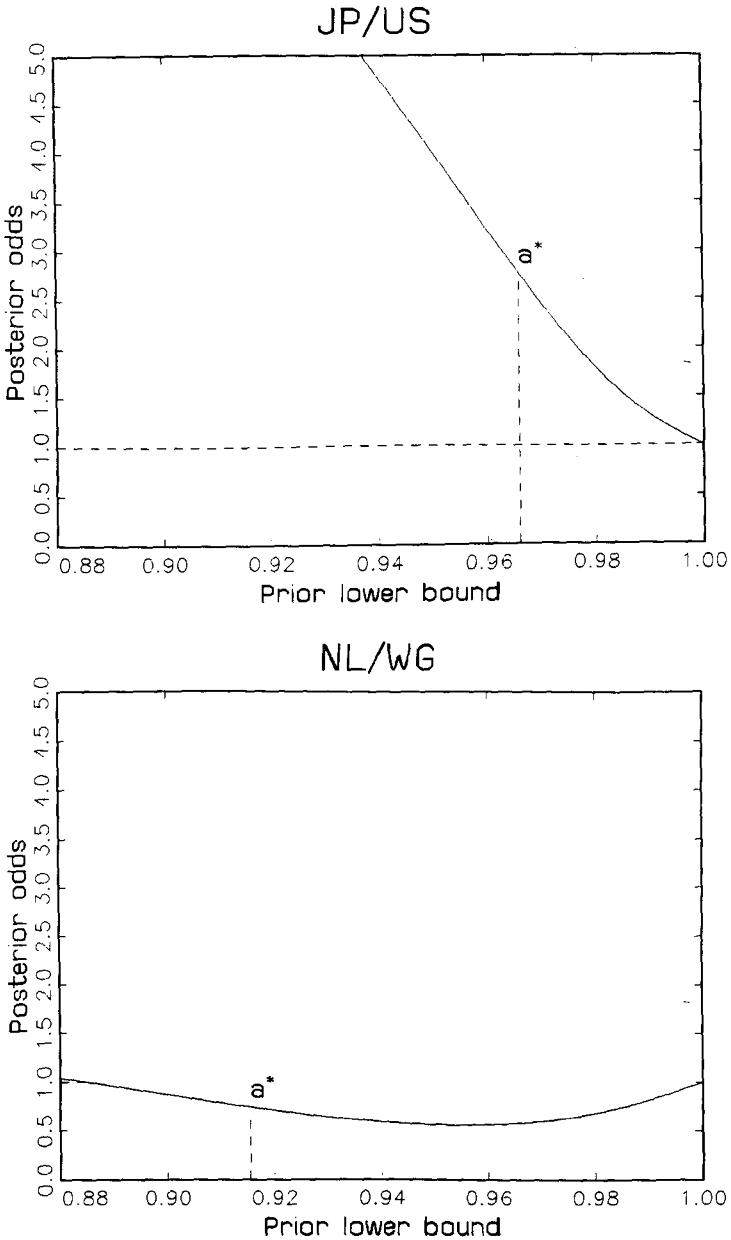


Fig. 6 (continued)

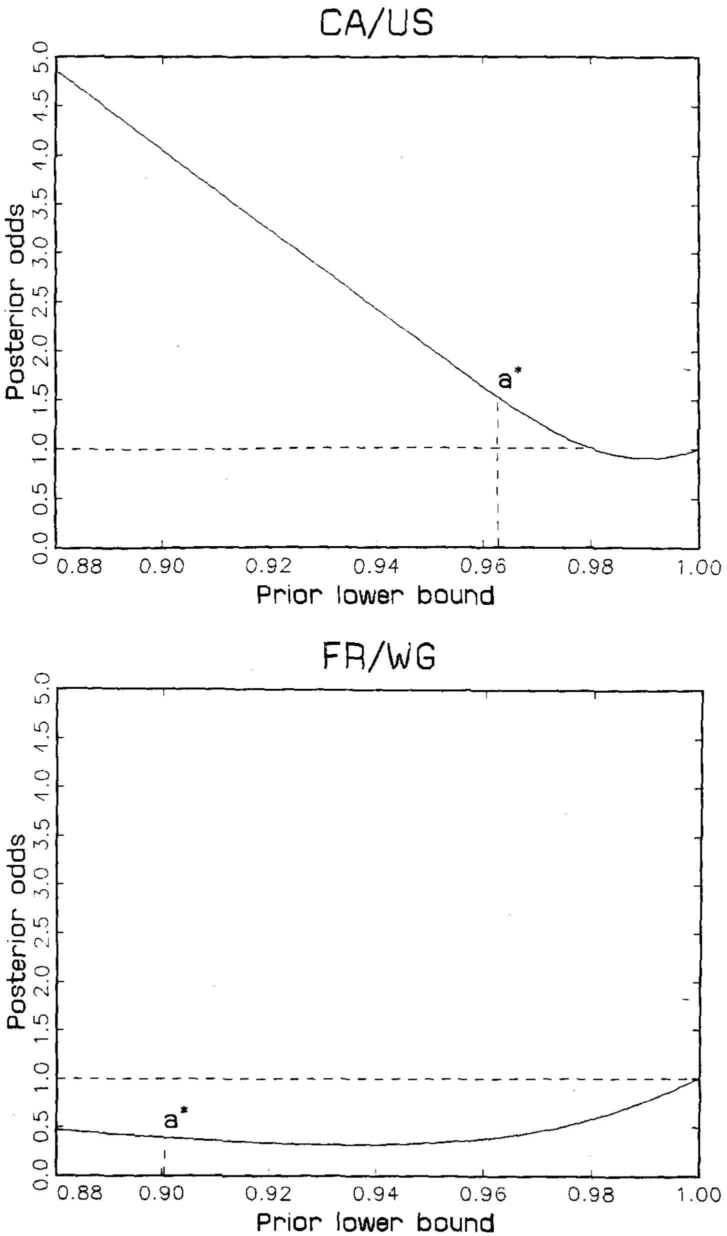


Fig. 6 (continued)

to find that the posterior is actually very concentrated near high values of ρ, leading him to revise his prior quite strongly and rejecting stationarity.

The prior regions that lead to stationarity are not unreasonable given previous research. Based on very long time series Frankel and Meese (1987) argue that the adjustment speed of real exchange rates is approximately 20% a year. This implies that the first-order autocorrelation with annual data is expected to be about $1/1.2 = 0.83$, and on monthly data $0.83^{1/12} = 0.985$. With the prior mean set to 0.985 the lower bound will be 0.97. For this value of a the odds are slightly in favour of stationarity, except for JP/US and CA/US.

5. Concluding remarks

We have performed a posterior odds analysis in order to compare a first-order autoregressive model containing a unit root (a so-called random walk model) with a first-order stationary autoregressive model. The effect of the presence of a constant term in the model, which leads to an identification problem under the random walk hypothesis and to a nonlinear estimation problem under the stationary alternative, has been studied. We have given special attention to the specification of reasonable prior distributions that are not in conflict with the data. Empirical results on time series of real exchange rates indicate that a Bayesian analysis can lead to different conclusions concerning the random walk behaviour of real exchange rates.

The results of the posterior odds lead to a different interpretation of the classical tests. Most outcomes neither strongly favour the unit root, nor the stationary AR(1) model. But adopting a classical testing procedure one would never have rejected the random walk. The classical test procedure, taking the random walk as the null hypothesis, only emphasizes that the random walk might be an appropriate model for real exchange rates. It fails to recognize, however, that there is a good alternative. A stationary AR(1) can describe the data about equally well.

The test procedures discussed in the paper can easily be extended to higher order autoregressive dynamics as considered in DeJong and Whiteman (1989). In our setup the additional parameters appear symmetrically under the unit root hypothesis and under the stationary alternative. They can therefore be treated as nuisance parameters. This means that we can specify uninformative priors for these parameters without running into the complications that arise from the treatment of the constant term. Likewise the extension to models with a nonzero growth rate will also not greatly complicate the analysis.

Diagnostic testing indicated that some of the assumptions, such as normality, are clearly violated. However, with Sims (1988) we conjecture that our results are not very sensitive to the normality assumption. In fact, we note that the nonnormality is mainly due to some outliers. In general we suspect that these phenomena are unavoidable in a mechanical time series modelling approach where the effects of policy interventions, like an EMS realignment, are not specified. Modelling properly anticipated policies is difficult, and clearly beyond the scope of this paper.

Appendix: Posterior odds for AR(1) with constant term

Proof of the theorem in section 3.2

Let $f(\mu,\rho)=(u'u)^{-T/2}$, with $u'u$ as defined in (16), be the kernel of the marginal posterior of μ and ρ.[10] To compute the posterior odds we need the integral $\int_{-M/2}^{M/2} f(\mu,1)\,\mathrm{d}\mu$ and the double integral $\int_{-M/2}^{M/2}\int_a^1 f(\mu,\rho)\,\mathrm{d}\rho\,\mathrm{d}\mu$, and then divide them by M and $M(1-a)$, respectively. Clearly the function $(u'u)^{-T/2}$ has an upper bound, since the quadratic form $u'u$ can not become zero for any μ and ρ, and it has the lower bound zero. As a consequence the posterior is integrable on a bounded region, where $-M/2<\mu<M/2$ for any finite $M>0$. Whether it is integrable when $M\to\infty$ remains to be investigated. In particular, we will study the behaviour of the posterior odds when $M\to\infty$.

Since $f(\mu,1)$ does not depend on μ, the first integral is simply $M(\Delta y'\Delta y)^{-T/2}$. To evaluate the double integral we first integrate over ρ, using the properties of the Student-t. The result is

$$f(\mu)=\int_a^1 f(\mu,\rho)\,\mathrm{d}\rho$$

$$\propto S^2(\mu)^{-T/2}\cdot s_{\hat\rho(\mu)}$$

$$\cdot\left[F\left(\frac{1-\hat\rho(\mu)}{s_{\hat\rho(\mu)}};T-1\right)-F\left(\frac{a-\hat\rho(\mu)}{s_{\hat\rho(\mu)}};T-1\right)\right], \qquad \text{(A.1)}$$

where

$$\hat\rho(\mu)=\frac{(y-\iota\mu)'(y_{-1}-\iota\mu)}{(y_{-1}-\iota\mu)'\,(y_{-1}-\iota\mu)},$$

$$S^2(\mu)=((y-\iota\mu)-\hat\rho(\mu)(y_{-1}-\iota\mu))'$$
$$\times((y-\iota\mu)-\hat\rho(\mu)(y_{-1}-\iota\mu)),$$

$$s^2_{\hat\rho(\mu)}=S^2(\mu)((y_{-1}-\iota\mu)'(y_{-1}-\iota\mu))^{-1}.$$

[10] For convenience we omit the conditioning argument on the data Y.

The proportionality sign is used to suppress integrating constants that do not depend on μ. If $\mu \to \infty$, then $\hat{\rho}(\mu) \to 1$, $S^2(\mu) \to (\Delta y'\Delta y - (y_T - y_0)^2/T)$, and $s^2_{\hat{\rho}(\mu)} \to 0$. Further, $(1-\hat{\rho}(\mu))/s_{\hat{\rho}(\mu)}$ converges to some finite constant depending only on the data. Hence the term between brackets involving the truncated t-distribution converges to some positive constant. The function $f(\mu)$ is therefore bounded by $ks_{\hat{\rho}(\mu)}$ for $\mu > \bar{\mu} > 0$ and some finite $k > 0$. Since $s_{\hat{\rho}(\mu)}$ is of the order $1/\mu$, the integral $\int_{\bar{\mu}}^{M/2} f(\mu)\,\mathrm{d}\mu$ will be of the order $\ln M$ or less. Finally, for the posterior odds we need the averaged likelihood under the null and the alternative, which involves dividing the numerator by M, and the denominator by $M(1-a)$. Recalling that the numerator $\int_{-M/2}^{M/2} f(\mu,1)\,\mathrm{d}\mu$ was of the order M, the net result is that the odds are of the order $M/\ln M$. The conclusion is that the odds will go to infinity if $M \to \infty$.

Computation of posterior odds with normal prior on μ

The complete specification of the prior has been discussed in sections 2 and 3 in the text. Combining prior and likelihood we obtain the posterior

$$p(\mu,\rho,\sigma|Y)$$

$$\propto \vartheta(2\pi)^{-T/2}\sigma^{-(T+1)}\exp\left(-\frac{1}{2\sigma^2}\Delta y'\Delta y\right)g(\mu) \qquad \text{if } \rho = 1, \tag{A.2}$$

$$\propto \frac{1-\vartheta}{1-a}(2\pi)^{-(T+1)/2}\sigma^{-(T+2)}\exp\left(-\frac{1}{2\sigma^2}\left(u'u + (1-\rho^2)(y_0-\mu)^2\right)\right) \qquad \text{if } \rho \in S,$$

where $u'u = (y - \iota\mu(1-\rho) - \rho y_{-1})'(y - \iota\mu(1-\rho) - \rho y_{-1})$ is the residual sum of squares as a function of μ and ρ. The constants of proportionality in (A.2) are the same for $\rho = 1$ as for $\rho \in S$. To compute posterior odds we first marginalize the posterior over σ and μ.

In the case $\rho = 1$ we have, after integrating over σ, that

$$\Pr(\rho = 1|Y) \propto \vartheta\Gamma(T/2)(\Delta y'\Delta y)^{-T/2}. \tag{A.3}$$

Note that $g(\mu)$, the prior on μ, is irrelevant, since μ does not appear in the likelihood when $\rho = 1$.

In the stationary case we also first integrate over σ, obtaining

$$p(\mu, \rho | Y) \propto \frac{(1-\vartheta)}{(1-a)} \pi^{-1/2} \Gamma\left(\frac{T+1}{2}\right)(1-\rho^2)^{1/2}$$

$$\times \left(u'u + (1-\rho^2)(\mu - y_0)^2\right)^{-(T+1)/2}. \tag{A.4}$$

To integrate over μ we rewrite the quadratic forms in the last factor in (A.4) as

$$u'u + (1-\rho^2)(\mu - y_0)^2$$

$$= \left(T(1-\rho)^2 + 1 - \rho^2\right)(\mu - \hat{\mu}(\rho))^2 + S^2(\rho)$$

$$= S^2(\rho)\left[1 + \frac{1}{T}\left(\frac{\mu - \hat{\mu}(\rho)}{s_{\hat{\mu}(\rho)}}\right)^2\right], \tag{A.5}$$

where

$$\hat{\mu}(\rho) = \frac{\iota' y - \rho \iota' y_{-1} + (1+\rho) y_0}{T(1-\rho) + 1 + \rho},$$

$$S^2(\rho) = e'e + (1-\rho^2)(y_0 - \hat{\mu}(\rho))^2,$$

$$e = y - \iota\hat{\mu}(\rho)(1-\rho) - \rho y_{-1},$$

$$s^2_{\hat{\mu}(\rho)} = \frac{S^2(\rho)}{T}\left(T(1-\rho)^2 + 1 - \rho^2\right)^{-1}.$$

$S^2(\rho)$ is a residual sum of squares. From (A.5) it follows that μ has a t-distribution conditional on ρ. Hence we can analytically integrate (A.4) over μ, using the integration formula

$$\int_{-\infty}^{\infty}\left[1 + \frac{1}{T}\left(\frac{\mu - \hat{\mu}(\rho)}{s_{\hat{\mu}(\rho)}}\right)^2\right]^{-(T+1)/2} \mathrm{d}\mu = s_{\hat{\mu}(\rho)} \frac{\Gamma(1/2)\Gamma(T/2)}{\Gamma((T+1)/2)}\sqrt{T}. \tag{A.6}$$

Hence the marginal posterior of ρ becomes

$$p(\rho|Y) \propto \frac{(1-\vartheta)}{(1-a)}\Gamma(T/2)(1-\rho^2)^{1/2} \times \left(T(1-\rho)^2+1-\rho^2\right)^{-1/2} S^2(\rho)^{-T/2}. \tag{A.7}$$

Since $\hat{\mu}(\rho)$ is a nonlinear function of ρ, $S^2(\rho)$ cannot be written as a quadratic form in ρ. Hence the marginal posterior density is not of the Student-t type. The marginal posterior also contains the factor

$$h(\rho) = (1-\rho^2)^{1/2}\left(T(1-\rho)^2+1-\rho^2\right)^{-1/2} = \left(1+T\frac{1-\rho}{1+\rho}\right)^{-1/2}, \tag{A.8}$$

which is of the order $1/\sqrt{T}$ for $\rho \neq 1$, and equals one for $\rho = 1$. Since both $S(\rho)$ and $h(\rho)$ are bounded functions of ρ, the marginal posterior is integrable. The integral has to be evaluated numerically, though. The posterior odds are obtained as the ratio of (A.3) to the average of (A.7):

$$K_1 = \frac{\vartheta}{1-\vartheta} \cdot \frac{(\Delta y'\Delta y)^{-T/2}}{\frac{1}{1-a}\int_a^1 h(\rho)S(\rho)^{-T/2}\,\mathrm{d}\rho}. \tag{A.9}$$

References

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

Bhargava, A., 1986, On the theory of testing for unit roots in observed time series, Review of Economic Studies 53, 369–384.

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

Cochrane, J.H., 1988, How big is the random walk in GNP?, Journal of Political Economy 96, 893–920.

Christiano, L.J. and M. Eichenbaum, 1990, Unit roots in real GNP: Do we know and do we care?, Carnegie–Rochester Conference Series on Public Policy 32, 7–62.

DeJong, D.N. and C.H. Whiteman, 1989, Trends and random walks in macroeconomic time series: A reconsideration based on the likelihood principle, Working paper (University of Iowa, Iowa City, IA).

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

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

Diebold, F.X., 1988, Empirical modeling of exchange rate dynamics (Springer Verlag, New York, NY).

Dornbusch, R., 1976, Expectations and exchange rate dynamics, Journal of Political Economy 84, 1161–1176.
Dornbusch, R. and J.A. Frankel 1987, The flexible exchange rate system: Experience and alternatives, NBER working paper 2464.
Edison, H.J. and E.O. Fisher 1988, A long-run view of the European Monetary System, International finance discussion paper XXX (Federal Reserve Board, Washington, DC).
Evans, G.B.A. and N.E. Savin, 1984, Testing for unit roots: 2, Econometrica 52, 1241–1269.
Frankel, J.A., 1979, On the mark: A theory of floating exchange rates based on real interest differentials, American Economic Review 69, 610–622.
Frankel, J.A. and R. Meese 1987, Are exchange rates excessively variable?, NBER Macroeconomics Annual.
Fuller, W.A., 1976, Introduction to statistical time series (Wiley, New York, NY).
Huizinga, J., 1987, An empirical investigation of the long run behavior of real exchange rates, Carnegie-Rochester Conference Series on Public Policy 27, 149–214.
Kaminsky, G., 1988, The real exchange rate in the short and in the long run, Working paper 88-11 (University of California, San Diego, CA).
Leamer, E.E., 1978, Specification searches (Wiley, New York, NY).
Mecagni, M. and P. Pauly, 1987, Relative prices and exchange rates in the long-run, Unpublished working paper (University of Pennsylvania, Philadelphia, PA).
Meese, R.A. and K. Rogoff, 1983a, Empirical exchange rate models for the seventies: Do they fit out-of-sample?, Journal of International Economics 14, 3–24.
Meese, R.A. and K. Rogoff, 1983b, The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification?, in: J.A. Frenkel, ed., Exchange rates and international macroeconomics (University of Chicago Press, Chicago, IL).
Meese, R.A. and K. Rogoff, 1988, Was it real? The exchange rate–interest differential relation over the modern floating rate period, Journal of Finance 43, 933–948.
Perron, P., 1988, Trends and random walks in macroeconomic time series: Further evidence from a new approach, Journal of Economic Dynamics and Control 12, 297–332.
Phillips, P.C.B., 1987, Time series regression with a unit root, Econometrica 55, 277–302.
Sims, C.A., 1988, Bayesian skepticism on unit root econometrics, Journal of Economic Dynamics and Control 12, 463–474.
Spanos, A., 1986, Statistical foundations of econometric modelling (Cambridge University Press, Cambridge).
Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York, NY).