ANNALS OF ECONOMICS AND FINANCE **3**, 43–64 (2002)

PPP May not Hold Afterall: A Further Investigation

Serena Ng

Department of Economics
Johns Hopkins University
E-mail: serena.ng@jhu.edu

and

Pierre Perron*

Department of Economics
Boston University
E-mail: perron@bu.edu

In a recent paper, Engel, C. (1999) presents monte-carlo evidence to suggest that unit root tests cannot detect a non-stationary component in the real exchange rate even when this component accounts for almost half of its long-horizon forecast error variance. This hidden non-stationary component led to the conclusion that long run purchasing power parity might not hold afterall. In this note, we first point out some conceptual difficulties with the statistic being used to measure the size of the non-stationary component, and then argue that it bears no systematic relationship with rejection rates in unit root tests. The problems stem from near observational equivalence of the simulated model in not one, but two dimensions. We then discuss the steps a practitioner can take to minimize Type I error in cases when the non-stationary component is hard to detect. Real exchange rate data for 19 countries are examined and estimates are obtained for the duration of the real exchange rate shocks.

Key Words: PPP; Monte-carlo; Unit root.
JEL Classification Numbers: F30, F40, C2, C3, C5.

* We thank Charles Engel for encouraging us to pursue this research. Correspondence: Serena Ng, Department of Economics, Johns Hopkins University, 3400 N. Charles st. Baltimore, M.D, 21218

1529-7373/2002

1. MOTIVATION

The law of one price or purchasing power parity (PPP) posits that the price level of traded goods converted to a common currency should be equal as a result of arbitrage. It would then seem natural to conjecture that national price levels converted to a common currency, the real exchange rate, should also tend towards parity. Let $p = (1-\psi)p^T + \psi p^N$ be the logarithm of the national price level of a home country. It is a geometric weighted average of the log price of traded, p^T, and non-traded goods, p^N, where ψ is the share of non-traded goods. If we denote variables for the foreign country with an asterik (*), and let s_t be the log of the nominal exchange rate, then the log real exchange rate is

$$\begin{aligned} q_t &= s_t + p_t^* - p_t \\ &= x_t + y_t, \end{aligned} \tag{1}$$

where

$$\begin{aligned} x_t &= s_t + p_t^{T*} - p_t^T, \\ y_t &= \psi^*[p_t^{N*} - p_t^{T*}] - \psi[p_t^N - p_t^T]. \end{aligned}$$

The real exchange rate thus has two components: a traded-goods component x_t, and a component y_t which captures the bilateral difference between the relative price of traded to non-traded goods. The real exchange rate is stationary if x_t and y_t are both stationary, or x_t and y_t are both non-stationary but that the two series are cointegrated in a particular way. Given our prior that PPP should hold for traded goods, at least in the long run, non-stationarity of x_t is difficult to admit. The stationarity of the real exchange rate would then seem to rest on the stationarity of y_t. Is this necessary? And does it matter if y_t is non-stationary?

Real exchange rates are generally found to be highly persistent. There is little dispute about this. More difficult to ascertain is whether this persistence is strong enough to be deemed non-stationary. Establishing this as a fact turns out to be a non-trivial task. As surveyed in Froot, K. and K. Rogoff (1995) while early research suggests the presence of a unit root, the recent evidence supports stationarity. Many put the blame for this ambiguity on the low power of unit root tests and the lack of data with a long enough span. The concerns expressed in Froot, K. and K. Rogoff (1995) are representative of the latter view. The size of unit root tests was rarely brought up as an issue.

In a recent paper, Engel, C. (1999) argued that unit root tests would fail to identify a non-stationary component in the real exchange rate even if there was one. He used a macroeconomic model with eight parameters to calibrate the U.S.-U.K. real exchange rate. The data support x_t as a stationary $AR(1)$ process but suggest that y_t is integrated, implying that $q_t = x_t + y_t$ should also be integrated. But in monte carlo experiments in which q_t was tested for a unit root using standard methods, unit root tests would overwhelmingly reject the null hypothesis even though q_t has a unit root component by construction, and that the long horizon forecast error variance of y_t accounts for almost half of the combined forecast error variance of x_t and y_t. Rejection rates were close to 100 percent when asymptotic critical values at the 5% level were used. So why do unit root tests reject non-stationarity?

The goal of this article is to help understand some of these results. We also use the PPP example to explain in a non-technical manner why testing for a unit root in some data is quite difficult and to stress that in spite of this difficulty, one can improve upon standard methods, both in terms of minimizing Type I error (rejecting the null hypothesis of non-stationarity when true) and maximizing the power of the tests. We analyze data on real exchange rates and present evidence for the half life of shocks to real exchange rates using estimates of autoregressive parameters that have better properties than those obtained from conventional methods.

The rest of this paper is structured as follows. Section 2 discusses theoretical issues concerning the presence of a negative moving-average component in real exchange rate series. Section 3 outlines potential problems that standard tests for unit root face in such cases and considers an improved methodology based on Ng, S. and P. Perron (2001). Section 4 illustrates the relevant issues with exchange rate data and Section 5 offers some concluding comments.

2. THE NEGATIVE MOVING AVERAGE COMPONENT AND THE REAL EXCHANGE RATE

Consider a series $\tilde{z}_t = z_t - \mu_t$, $(t = 1, \ldots T)$ where μ_t is a deterministic trend function which we assume is known for the moment. Suppose the (demeaned or detrended) series $\tilde{z}_t$ is generated by an unobserved component model, with $\eta_t \sim i.i.d.(0, \sigma_\eta^2)$ and $v_t \sim i.i.d.(0, \sigma_v^2)$:

$$\begin{aligned} \tilde{z}_t &= \tau_t + \eta_t, \\ \tau_t &= \alpha\tau_{t-1} + v_t. \end{aligned}$$

Then

$$\Delta\tilde{z}_t = (\alpha - 1)\tilde{z}_{t-1} + u_t, \qquad (2)$$

where

$$u_t = e_t + \theta e_{t-1},$$

with $e_t \sim i.i.d.(0, \sigma_e^2)$ and θ satisfying

$$\frac{\theta}{1+\theta^2} = \frac{-\alpha\sigma_\eta^2}{\sigma_v^2 + (1+\alpha^2)\sigma_\eta^2}.$$

Suppose $\alpha = 1$ and σ_v^2 (the variance of the innovation to τ_t) is infinite; then $\theta = 0$ and $\tilde{z}_t$ is completely dominated by the random walk component. At the other extreme when $\sigma_v^2 = 0$, $\theta = -1$ and $\tilde{z}_t = u_t$ is *i.i.d.* in view of the common factor between the moving average and autoregressive polynomials (the unit root). In between the two extremes, $\tilde{z}_t$ is fundamentally non-stationary but has a tendency to revert to mean. This force for mean reversion is larger the closer is θ to -1. It is this tension between non-stationarity and mean reversion that poses problems for unit root tests. The size problem arises because $\tilde{z}_t$ behaves much like a stationary process. Cases when the innovation variances are of comparable magnitudes but α is near but not exactly unity are also a problem for unit root tests. But the problem there is low power and the issue should be kept distinct from the size problem arising from a near common factor in the moving-average and autoregressive polynomials that is being discussed here.

The size problem in testing for a unit root when there is a large negative moving average component was documented in Phillips, P. C. B. and P. Perron (1988) and highlighted by Schwert, G. W. (1989), among many others. To see the nature of the problem, rewrite (2) in the form of a $k^{th}-$order augmented autoregression in $\Delta\tilde{z}_t$:

$$\begin{aligned} \Delta\tilde{z}_t &= (\alpha-1)\sum_{i=0}^{k}(-\theta)^i \tilde{z}_{t-i-1} - \sum_{i=1}^{k}(-\theta)^i \Delta\tilde{z}_{t-i} + e_t - (-\theta)^{k+1}e_{t-k-1}. \\ &= \beta_0\tilde{z}_{t-1} + \sum_{i=1}^{k}\beta_i\Delta\tilde{z}_{t-i} + e_{tk}, \qquad (3) \end{aligned}$$

where $\beta_0 = (\alpha - 1)$, $\beta_i = -(-\theta)^i$. Notice that the truncation lag k plays a crucial role in the dynamic properties of $\Delta\tilde{z}_t$. When θ is large and negative, lags of $\Delta\tilde{z}_t$ will have non-negligible weights at large k. If $\theta = -.8$, for example, we need k to be at least 20 for $(-\theta)^k$ to be less than .01. Because $\Delta\tilde{z}_t$ is serially correlated, e_{tk} can be strongly serially correlated if k is small and $\theta < 0$. The severity of this problem is specific to negative values of θ because when θ is positive, $(-\theta)^i$ alternates in sign and successive lags of $\Delta\tilde{z}_{t-i}$ offset each other.

The size problem in (perhaps all) unit root tests when θ is negative can be traced to the fact that β_0 cannot be precisely estimated from (3). Nabeya, S. and P. Perron (1994) and Perron, P. (1996) analyzed the problem for the case with $k = 0$. The more general case which allows k to increase with the sample size was analyzed in Ng, S. and P. Perron (1995, 1997, 2001) and Perron, P. and S. Ng (1996, 1998). In this case, $\beta_0 + 1 \equiv \alpha$ is the sum of the coefficients of an $AR(k+1)$ model in the levels of $\tilde{z}_t$; it is this sum that is not precisely estimated.

Let us return to the real exchange rate problem. Suppose

$$\begin{aligned} y_t &= y_{t-1} + w_t, \\ x_t &= \phi x_{t-1} + m_t \end{aligned}$$

where w_t and m_t are $i.i.d.$ with variance σ_w^2 and σ_m^2, respectively, and covariance $\sigma_{wm}^2 \neq 0$. Engel, C. (1999) considered a three equations model for exchange rate determination (reproduced in the Appendix) for which x_t and y_t have the above time series properties and estimated the model using quarterly U.S./U.K. data over the sample 1970-1995. The estimates are then used to simulate 400 data points to mimic a 100 years sample, and a battery of unit root tests was constructed. When estimation of an autoregression such as (3) was required, k was set to a maximum (hereafter denoted $kmax$) of 12 and a chi-square test was then used to test for the significance of the last lag.[1] Huge size distortions in unit root tests for both the baseline parameters and for small perturbations around them were found. The MZ_α test[2] developed in Perron, P. and S. Ng (1996) to be more robust to size distortions when θ is negative did not work as it should. In line with these results, tests for the null hypothesis of stationarity also support stationarity. Engel also evaluated, as a measure of the importance of the random walk component at horizon h, the function

$$\begin{aligned} R_0(h) &= \frac{mse(y_{T+h} - E[y_{T+h}|x_t, y_t, t \leq T])}{mse(q_{T+h} - E[q_{T+h}|x_t, y_t, t \leq T])} \qquad (4) \\ &= \frac{h \cdot \sigma_w^2}{h \cdot \sigma_w^2 + \frac{(1-\phi^{2h})}{(1-\phi^2)}\sigma_m^2 + 2\frac{(1-\phi^h)}{(1-\phi)}\sigma_{wm}^2} \end{aligned}$$

where $mse(\cdot)$ denotes the mean-squared forecast error function and h is the forecast horizon.[3] Assuming σ_{wm}^2 is negligible, this measure can be approximated by

$$R_0(h) = \frac{mse(y_{T+h} - E[y_{T+h}|y_t, t \leq T])}{mse(x_{T+h} - E[x_{T+h}|x_t, t \leq T]) + mse(y_{T+h} - E[y_{T+h}|y_t, t \leq T])}.$$

In this case, $R_0(h)$ is interpreted as the relative variance of the random walk component. For the stochastic processes assumed for x_t and y_t, and with

[1] This is a small variation to the t test considered in Ng, S. and P. Perron (1995).

[2] C. Engel referred to this as the PN test.

[3] Under optimal prediction, this is simply the forecast error variance and these terminologies will be used interchangeably.

$h = 400$ (a hundred years horizon), R_0 was reported to be around .4 for the base case. At face value, this implies that variations in the non-stationary component are important at long horizons. Yet, unit root tests reject non-stationarity and tests for stationarity cannot reject that null hypothesis. It appears that unit root tests have indeed missed a non-negligible permanent component badly.

Recall from the unobserved components model that the problem of a small innovation variance in the random walk component maps into a large negative moving-average component in the observed series.[4] If y_t is a random walk and x_t is a stationary $AR(1)$, then for $q_t = x_t + y_t$,

$$\begin{aligned} \Delta q_t &= \phi \Delta q_{t-1} + e_t \\ e_t &= u_t + \theta u_{t-1} \end{aligned} \qquad (5)$$

There are two autoregressive roots in the model for q_t, 1, and ϕ, and it is a special case of (2) when $\phi = 0$. The key parameter is once again θ and it is related to the parameters of the processes for x_t and y_t as follows:

$$\frac{\theta}{1+\theta^2} = \frac{-\phi\sigma_w^2 - \sigma_m^2 - (1+\phi)\sigma_{wm}^2}{(1+\phi^2)\sigma_w^2 + 2\sigma_m^2 + 2(1+\phi)\sigma_{wm}^2}. \qquad (6)$$

For the base case analyzed by Engel, C. (1999), $\sigma_w^2 = .328 \times 10^{-4}$ and $\sigma_m^2 = .2667 \times 10^{-2}$ with σ_{wm}^2 very small. Since σ_w^2 is 100 times smaller than σ_m^2, θ should be large and negative. For the base case, Engel reported a value of $\theta = -.8$ and argued that unit root tests fail to detect a unit root because this negative moving average component induces size distortions. Our concerns are two fold. One is about using R_0 as a measure of the non-stationary component, and the other is relating R_0 to the size of unit root tests.

To understand the relevant issues, we first take three configuration of the parameters and repeat Engel's monte-carlo exercise.[5] The first three rows of Table 1 report results from 2500 replications and the rejection rates of unit root tests are indeed large. Table 1 also confirms that R_0 is about .4 in the base case. But consider the values of ϕ and θ implied by configurations of the exchange rate model (given in the Appendix). Observe that θ in Case 1 (the base case) is $-.9911$, not $-.8$ that Engel reported. For cases 2 and 3, θ is $-.9995$ and $-.9827$, respectively. The common factor problem is thus more severe than Engel thought. This extreme form of parameter redundancy in the simulated data will be important in understanding subsequent results.

[4] Ng, S. and P. Perron (1997) used a similar framework to analyze the inflation series.
[5] Case 1 is the base case of Engel. Other configurations in his Table 3 give very similar parameter values and therefore have similar size properties.

TABLE 1.

Statistics with k selected by the t test, $kmax = 12$.

Case	ϕ	θ	MZ_α	DF	Z_α	$\hat{\alpha}$	$R_0(400)$
1	0.919	-0.991	0.960	0.912	0.966	0.918	0.431
2	0.976	-0.999	0.443	0.304	0.459	0.965	0.009
3	0.952	-0.982	0.474	0.362	0.485	0.962	0.840
4	0.919	-0.925	0.083	0.077	0.088	0.984	0.995
5	0.850	-0.892	0.154	0.130	0.164	0.976	0.990
6	0.850	-0.971	0.847	0.800	0.870	0.894	0.806
7	0.999	-0.999	0.072	0.057	0.076	0.985	0.017
8	0.839	-0.982	0.966	0.948	0.977	0.859	0.594

Notes: ϕ and θ are the coefficients corresponding to the model $\Delta q_t = \phi\Delta q_{t-1}+e_t$ and $e_t = u_t+\theta u_{t-1}$ as implied by the parameters in Table A.1 in the Appendix. The entries under MZ_α, DF, and Z_α show the size of the tests (the rejection rates are based on the 5% asymptotic critical values of -14.1, -2.86, and -14.1). The tests are based on OLS demeaned data. The entries under $\hat{\alpha}$ are the average of the statistics over the replications. $R_0(400)$ is defined in (4).

More fundamentally, R_0 is not a satisfactory measure of the relative size of the non-stationary component for two reasons. First, as a measure of the relative importance of the non-stationary component in a series, one would expect the statistic to be bounded between zero and one. It is easy to see that R_0 fails this criterion for some choices (though not those considered by C. Engel) of the variances and covariances. As well, the statistic is discontinuous at $\phi = 1$ and unstable when ϕ is near the unit circle. Third, by definition, the random walk component should eventually explain 100% of the forecast error variance. Thus, when using R_0 to infer the size of the random walk component, account must also be taken of the duration required to explain x% of the forecast error variance. For the base case, R_0 is is still below unity after 2,000 periods, or 500 years. The impression being portrayed that R_0 is .4 after 100 years and thus has a non-trivial random walk component is in some sense misleading.

The more serious problem stems from extending the intuition that unit root tests fail to detect the random walk component because it is small, to presuming a systematic inverse relation between rejection rates in unit root tests and R_0. To see that this relation does not hold up, we consider other parameterizations of the model. These are labelled Cases 4 through 8 in Table 1. Case 4 has a large R_0 and yet rejection rates of unit root tests are small. Case 4 and Case 5 have similar values for R_0, and yet unit root tests have rather different rejection rates. Case 4 has R_0 similar to Case 3 but the rejection rates are much higher. Case 7 has a smaller R_0 than the case 1, and yet there is no size distortion. R_0 is larger in Case 8 than the base case, and yet unit root tests reject at least as often as in the base case.

The reason why R_0 bears no systematic relation to the rejection rates is that unit root tests are based on the properties of q_t, but R_0 is based on the components of q_t. Although q_t is linearly related to x_t and y_t, the dynamic properties of q_t are related to the parameters of x_t and y_t in a non-linear way. Forecasting q_t from linear combinations of the history of x_t and y_t will generally be different from forecasting q_t directly from its own history. That is to say, $MSE(q_{T+h}|q_T)$ can be quite different from $MSE(q_{T+h}|x_T, y_T)$. Consider the simplest example with $\phi = 0$, $\sigma_m^2/\sigma_w^2 \to \infty$, so that q_t is really white noise with constant forecast variance at all horizons. Forecasts based on x_T and y_T will have mean-squared errors that resemble those of y_t and increase with h. Although it is generally not the case that forecasting the components can have mean-squared errors larger than forecasting the aggregate directly, when there are common factors the aggregate model which eliminates redundant information can potentially deliver more efficient forecasts. For the present exercise, the crucial point is that R_0 is most unreliable when there are common factors, but the size problem in unit root tests arises precisely because of those common factors.

For the model considered, $(1 - \phi L)\Delta q_t = (1 + \theta L)u_t$, three outcomes are possible: q_t can have one autoregressive root near unity, one root at exactly unity, and two autoregressive roots, one at unity and one near unity. The first scenario obtains if θ is -1 and cancels the unit root, so that q_t is an $AR(1)$ with parameter ϕ. Rejection rates in unit root tests then reflect the power of unit root tests. The second scenario obtains if θ cancels the second autoregressive root in which case q_t has a unit root. The third obtains if no cancellation occurs, so q_t has one root of unity and one root close to unity. Rejection rates in the latter two cases then reflect the size of the tests. What complicates the interpretation of the rejection rates here is a coefficient ϕ that is calibrated to be not too far from the unit circle and a moving average root that is close to both autoregressive roots.

The values of ϕ and θ provide a reasonable guide to the rejection rates reported in Table 1. Cases 4 and 7 both involve a mean common factor between ϕ and θ, but the unit root in the data is left intact. The rejections rates in Table 1 pertain to the size of the tests, which, as we can see, are close to the nominal size of 5%. Cases 2 also has a near moving average unit root that nearly cancels the unit root. Since the second autoregressive root is $.976 < 1$, the rejection rates resemble power. Rejection rates are low because ϕ lies in the parameter range for which unit root tests have low power. Cases 3, 5, 6, and 8 have values of θ somewhat further away from the unit circle, and for which we would expect unit root tests not to reject the null hypothesis. But size distortions remain noticeable. Thus, the size problem in unit root tests is genuine.

In cases of severe parameter redundancy, it could be argued that a precise I(1)/I(0) classification is quite meaningless or undesirable. A measure of persistence that is independent of such a clear-cut classification might be more appropriate. In this avenue, a useful concept relates to how fast the effects of shocks to q_t dissipate. It is thus of some interest to assess autoregression based measures of persistence in this setting. Let $\hat{\alpha} = \hat{\beta}_0 + 1$, obtained by applying OLS to (3), be the sum of the estimated autoregressive coefficients. Consider two statistics which aim to capture the time required for a fraction τ of the full effect of a unit shock to complete:

$$\begin{aligned} J_0^\tau &= \sup_j |\partial z_{t+j}/\partial u_t| \leq 1 - \tau, \\ J_1^\tau &= \log(1-\tau)/\log(\hat{\alpha}). \end{aligned}$$

When $\tau = .5$, J_0^τ is the period beyond which the (absolute) response to a unit shock in u_t no longer exceed .5. On the other hand, J_1^τ is the half life of a shock as implied by the estimated sum of the autoregressive coefficients. The difference between the two is that J_0^τ is based on the moving-average representation of the estimated model and, hence, depends on all parameters of the autoregressive representation. In contrast, J_1^τ depends only on the sum of the estimated autoregressive parameters.

Although both statistics will agree that complete adjustment to a shock will take infinitely long when there is a unit root, J_1^τ will overstate the duration of adjustment for intermediate values of τ (such as .5) when a series has a large negative moving average component. The reason is that when there is a negative moving average component, the coefficient associated with the moving-average representation at lag k can be much smaller than the sum of the k autoregressive coefficients.

More generally, conventional statistics of persistence should be interpreted with some care when the data are strongly mean reverting. Take the autocorrelation function, say, $\Gamma(j)$, which is also a widely used measure of persistence in macroeconomic analysis.[6] The potential problem with $\Gamma(j)$ is that it is typically evaluated at only small values of j. But for processes that are both persistent and have a tendency for mean reversion, the values of j needed when τ is high could be very large. Without considering a large j, $\Gamma(j)$ could understate how slow is the reversion to the mean. This issue is relevant when studying the persistence of series such as inflation.[7] For such data, the statistic J_0 has the advantage that we do

[6] For example, Mcgrattan, E., A. Chari, and P. Kehoe (1998) used the degree of serial correlation in the real exchange rate to judge whether a sticky price model can replicate the observed persistence in the real exchange rate. Bergin, P. and R. Feenstra (1999) used the first and fourth order autocorrelation coefficients to assess the degree of stickiness in the real exchange rate.

[7] Work is in progress to document this issue in more detail.

not need an a priori choice on j; it is endogenously determined once we pick the cut-off point, τ.

3. TESTING FOR A UNIT ROOT ONCE AGAIN

Some, including ourselves, have argued[8] that there is always a non-stationary representation for a time series that is arbitrarily close to a stationary representation. Because of this potential for observational equivalence, any test that has high power rejecting the null hypothesis of a unit root when the signal of the non-stationary component is strong must also have size distortions when this signal is weak. Consider once again the mapping from the relative magnitude of the innovation variance to θ. The near-observational equivalence problem can be stated as follows: when using unit root tests with asymptotic critical values, there will exist values of θ in the range $(-1, x)$ for some $-1 < x < 0$, say, such that liberal size distortions will surface. The value of x will depend on the sample size and the test used, but it will always approach -1 as the sample size increases. That is to say, the range over which size distortions occur should diminish.

The standard story about observational equivalence assumes the absence of a second autoregressive root that is large, and is not immediately applicable to the real exchange rate example in which ϕ is large. Nonetheless, in the simulated examples Engel considered, unit root tests really ought not to have rejected the null hypothesis of a unit root, though the defense that $\theta = -.99$ is not an interesting case could perhaps be invoked. The more serious problem is that for sample sizes commonly encountered, the value of x where size distortions start to appear is not $-.99$, but much further away from -1. Depending on the test, x could be anywhere from $-.4$ to $-.8$ for $T = 100$. This is worrisome because there will exist empirically important time series which are genuinely non-stationary, and would yet be classified as stationary. Cases 3, 5, 6 and 8 documented earlier are representative of the problem. Even with a sample size of 400, the statistics can reject with 80% probability instead of 5% for Case 6.

How prevalent are such time series? In our experience and as we will see in the next section, variables such as inflation tend to have this property, and we are in the process of a more complete documentation of such data. While a formal test of parameter redundancy is difficult because the maximum likelihood estimates of the autoregressive and moving average parameters are not precise when there is a near common factor,[9] the symptoms are there for us to detect. From our previous work, the kernel

[8]See Campbell, J. Y. and P. Perron (1991), Perron, P. and S. Ng (1996), Cochrane, J. H. (1991), Faust, J. (1996) and Blough, S. (1992).

[9]See Clark, P. K. (1988).

estimate of the spectral density at frequency zero based upon $\hat{e}_{t0}$ (i.e. the least squares residuals with no lagged first-differences included) should be very different from those based on $\hat{e}_{tk}$.[10] There should also be sharp differences between the Phillips-Perron Z tests and the MZ tests even though the two differ only by a term that should vanish at rate T. The premise of our latest work is precisely to exploit such information to robustify the size of the DF test and the class of MZ tests. This is achieved by parameterizing the model and/or finding estimators such that the sum of the autoregressive coefficients and the nuisance parameters can be estimated as precisely as possible. We now provide a non-technical summary of this work. All statistics considered are defined in the Appendix.[11]

To begin, recall that in the above discussion $\tilde{z}_t$ is the detrended series. That is, $\tilde{z}_t = z_t - \mu_t$, where μ_t is the deterministic component. For persistent data, least squares detrending is inefficient. Elliott, G., T. J. Rothenberg, and J. H. Stock (1996) showed that using GLS detrended data to construct the DF statistic can yield substantial power gains and Ng, S. and P. Perron (2001) showed that these power gains extend to the Z and MZ tests. As a first step, therefore, one should first quasi-transform the data at $\bar{\alpha} = 1 + \bar{c}/T$, where, as suggested by Elliott, G., T. J. Rothenberg, and J. H. Stock (1996), $\bar{c} = -7.0$ in the constant only case and -13.5 in the linear trend case. The estimates of the coefficients on the deterministic components are then obtained by OLS using the quasi differenced data. The discussion that follows assumes that all regressions are based on such GLS detrended data. We now discuss problems and suggest solutions for the DF^{GLS} and Z^{GLS} class of tests.

• The DF^{GLS} test. This test is the $t-$ statistic on β_0 in the k^{th} order augmented autoregression (3). The problem in the presence of a strong negative MA component is that $\hat{\beta}_0$ is severely biased if k is small because e_{tk} is serially correlated. The solution is then to select a large k when it is necessary. To implement this, and to avoid selecting a large k when not needed, we suggest using a modified information criterion, the $MAIC$ defined by, using (3) estimated from $t = kmax + 1, \ldots, T$ for all k,

$$MAIC = Argmin_{k=0,\ldots,kmax} \ln(\hat{\sigma}_k^2) + \frac{2(\tau_T(k) + k)}{T - kmax},$$

with

$$\tau_T(k) = (\hat{\sigma}_k^2)^{-1} \hat{\beta}_0^2 \sum_{t=kmax+1}^{T} \tilde{y}_{t-1}^2,$$

[10] These issues are discussed in Perron, P. and S. Ng (1998).

[11] This has been the basis of work reported in Ng, S. and P. Perron (1995, 1997, 2001), and Perron, P. and S. Ng (1996, 1998).

where $\hat{\sigma}_k^2 = T^{-1}\sum_{t=kmax+1}^{T}\hat{e}_{tk}^2$.

The key to the solution is an adequate selection of k, the order of the autoregression. The $MAIC$ is motivated by the observation that the bias in $\hat{\beta}_0$ decreases non-linearly as k increases. Model selection rules such as the AIC and BIC do not take this non-linearity into account; they under-penalize models with a small k and select autoregressive approximations that are too parsimonious for models with a negative MA component. The $MAIC$ explicitly accounts for the strong dependence of the bias in $\hat{\beta}_0$ on k via the term $\tau_T(k)$. The $MAIC$ reduces to the standard AIC when this dependence is absent (such as $ARMA$ noise functions with autoregressive and moving average roots far from the unit circle).

• The Z^{GLS} test. This test requires *(a),* a least squares estimate of α from the regression $\tilde{z}_t = \alpha\tilde{z}_{t-1} + e_{t0}$, and *(b)* estimates of the variance and the so-called long run variance (the non-normalized spectral density at frequency zero) of e_{t0}. The problem is that $\hat{\alpha}$ is severely biased because e_{t0} is strongly serially correlated and, because of this, estimates of both the variance and the long-run variance are also severely biased when constructed using the estimated residuals $\hat{e}_{t0}$. The solution is then to remove any dependence of Z^{GLS} on $\hat{\alpha}$. To implement this the following steps are taken. *(a)* Use the first differences of the data to construct an estimate of the variance. This leads to the so-called MZ^{GLS} test defined as

$$MZ_\alpha^{GLS} = \frac{T^{-1}\tilde{z}_T^2 - s^2}{2T^{-2}\sum_{t=1}^{T}\tilde{z}_{t-1}^2}.$$

where s^2 is the estimate of the long-run variance. *(b)* Estimate the long run variance using an autoregressive spectral density estimator based upon (3) with GLS detrended data. *(c)* When constructing the autoregressive spectral density estimate, use the $MAIC$ to select k.

Of the two tests, the MZ^{GLS} tests hold a size advantage while the DF^{GLS} has better power, especially for sample sizes less than 150.

It should be emphasized that proper implementation of the new tests is extremely important. Use of GLS detrending alone or the $MAIC$ alone will not reduce size distortions by as much. We still need to apply modifications to Z_α^{GLS}. Table 2 reports results for tests based on GLS detrended data, and for different values of $kmax$. Clearly, the size of Z_α^{GLS} is still inferior to MZ_α^{GLS}. Table 3 shows that, using the BIC to select k will still lead to size distortions, in spite of the use of GLS detrended data and the modified tests. The BIC selects $k = 1$ on average irrespective of $kmax$. With the $MAIC$, the average k is 12 when $kmax = 20$ and increases with $kmax$. Letting k be the default used in software packages is highly undesirable because the appropriate k is data dependent, and the model

TABLE 2.

Size of the Test Statistics with k selected by the MAIC.

	ϕ	θ	Z_α^{GLS}			MZ_α^{GLS}			DF^{GLS}		
Case/$kmax$			12	20	40	12	20	40	12	20	40
1	0.919	-0.991	0.706	0.586	0.369	0.692	0.556	0.342	0.682	0.546	0.308
2	0.976	-0.999	0.422	0.385	0.343	0.413	0.378	0.334	0.396	0.350	0.262
3	0.952	-0.982	0.337	0.280	0.206	0.329	0.268	0.197	0.309	0.244	0.145
4	0.919	-0.925	0.055	0.063	0.082	0.052	0.062	0.080	0.049	0.046	0.043
5	0.850	-0.892	0.087	0.074	0.092	0.084	0.071	0.088	0.078	0.061	0.052
6	0.850	-0.971	0.536	0.393	0.196	0.503	0.337	0.144	0.503	0.351	0.140
7	0.999	-0.999	0.063	0.067	0.078	0.059	0.066	0.078	0.054	0.054	0.036
8	0.839	-0.982	0.728	0.559	0.292	0.678	0.479	0.188	0.691	0.506	0.212

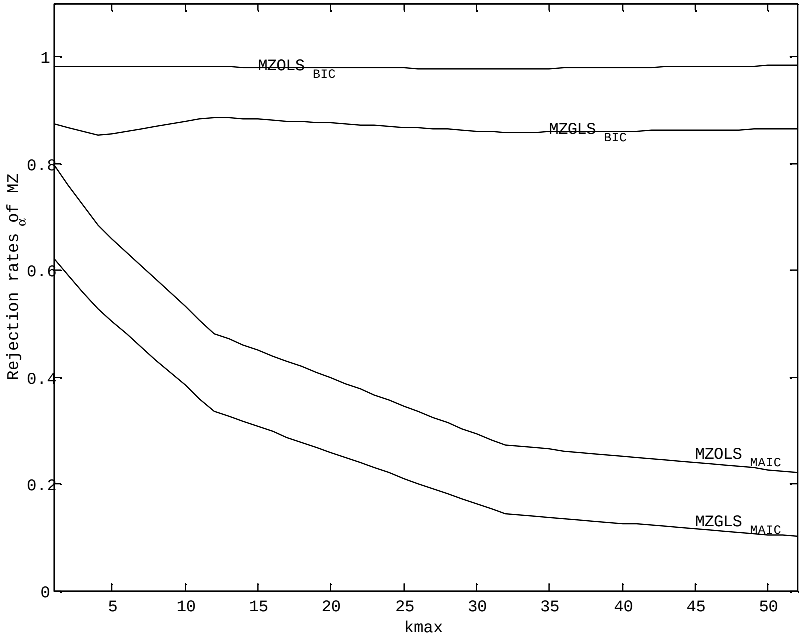
The 5% asymptotic critical value for Z_α is -14.1, for MZ_α^{GLS} is -8.1, and for DF^{GLS} is -1.91.

TABLE 3.

Size of the Test Statistics with k selected by the BIC.

	ϕ	θ	Z_α^{GLS}			MZ_α^{GLS}			DF^{GLS}		
Case/$kmax$			12	20	40	12	20	40	12	20	40
1	0.919	-0.991	0.889	0.882	0.888	0.887	0.880	0.884	0.888	0.882	0.886
2	0.976	-0.999	0.494	0.495	0.492	0.488	0.490	0.483	0.508	0.504	0.502
3	0.952	-0.982	0.445	0.427	0.422	0.441	0.424	0.418	0.451	0.433	0.422
4	0.919	-0.925	0.068	0.073	0.070	0.067	0.072	0.069	0.069	0.073	0.072
5	0.850	-0.892	0.162	0.168	0.180	0.160	0.164	0.176	0.162	0.170	0.176
6	0.850	-0.971	0.868	0.869	0.860	0.865	0.866	0.854	0.862	0.863	0.854
7	0.999	-0.999	0.070	0.067	0.060	0.067	0.066	0.057	0.068	0.066	0.061
8	0.839	-0.982	0.952	0.960	0.962	0.950	0.959	0.960	0.950	0.958	0.962

FIG. 1. Properties of MZ_α using the MAIC and the BIC: Case 6



selection rule must be flexible enough to handle the strongly mean reverting nature of data with negative moving average errors. Figure 1 uses case 6 to show how it is GLS detrending along with the selection of k that resolves the size problem. If we use OLS detrended data in the estimations and the BIC to select the lag length, the probability of rejecting the unit root hypothesis when it is true is close to unity. Replacing OLS by GLS detrending improves the size somewhat, as does replacing the BIC by the MAIC. However, it is the use of GLS detrending and the MAIC that yields large size improvement.

The only remaining issue is that, as with the AIC and the BIC, we need to specify a $kmax$ for the $MAIC$. Our theoretical results only provide guidance about the rate of increase of k relative to the sample size and does not pin down a $kmax$ for empirical work. Our recommendation is to use a $kmax$ that varies with the sample size, such as $kmax = Ont[12(T/100)^{1/4}]$. In our experience with simulations and applications, setting this $kmax$ is usually sufficient to obtain substantial size improvements. But because the k that is required to make e_{tk} approximately serially uncorrelated depends on the data generating process, there could be cases when such a $kmax$ might not be large enough. In that case, reset $kmax$ to some larger number and re-optimize the $MAIC$ function.

TABLE 4.

Output and Inflation in the U.S.

Series	$\hat{\alpha}_{ML}$	$\hat{\theta}_{ML}$	DF^{GLS}	DF	MZ_{α}^{GLS}	Z_{α}	$\hat{\alpha}^{GLS}$	$\hat{\alpha}^{OLS}$	$kmic$	$kbic$	$J_0^{.5}$	$J_1^{.5}$	$J_0^{.8}$	$J_1^{.8}$
log(gdp)	0.954	0.221	-1.576	-2.094	-5.319	-8.020	0.980	0.966	2	1	37	34	63	78
Inflation (gdp)	0.950	-0.298	-1.242	-2.834	-3.096	-15.388	0.967	0.895	2	0	14	20	56	48
Inflation (cpi)	0.947	-0.283	-1.199	-2.871	-3.089	-16.014	0.962	0.890	9	0	9	18	46	41

Regressions for log(GDP) include a constant and a trend. Only a constant is included in regressions for inflation.

3.1. Empirical Examples.

To illustrate, consider quarterly data for log(GDP) and the GDP and CPI based inflation series over 1962:1-1998:4 ($T = 160$)[12]. Both the DF^{GLS} and MZ^{GLS} use GLS detrending and the $MAIC$ with $kmax$ set to 14. We also report the DF and the Z_α tests, both based on least squares detrending and k selected using the BIC. The latter ones are perhaps the most commonly used methods in practice.

We first estimate an $ARMA(1,1)$ model to obtain a rough idea of the size of the moving average component. The results are reported in Table 4. For log(GDP), $\hat{\theta}$ is positive suggesting that, for all tests, size distortions are not an issue. Indeed, no test rejects a unit root in log(GDP) around a linear deterministic trend, and $MAIC$ and BIC chooses k at 2 and 1, respectively. Consider now the two inflation series. Both estimates of θ are negative, and even though the point estimates are far from -1, they are negative enough to cause problems. The Z_α test rejects a unit root in both cases, while the DF^{GLS} and MZ^{GLS} do not. The DF test is known to be somewhat more robust than Z_α when θ is negative, but it too rejects a unit root at the 5% level for one series and at the 10% level for the other. For both inflation series, the BIC selects a lag length of 0. The $MAIC$ selects 2 for the GDP-based series and 9 for the CPI-based series. This shows, first, that the BIC will not pick a large enough k, and second that the $MAIC$ does not necessarily pick the largest k possible. A comparison of $\hat{\alpha}^{GLS}$ and $\hat{\alpha}^{OLS}$ shows that in general, the OLS/BIC combination yields lower estimates of α.

Since a non-rejection of the unit root does not necessarily imply that this hypothesis is true, as discussed earlier it is useful to consider measures of the effects of shocks at various horizons such as the statistics J_0^τ and J_1^τ defined above. Evaluating them at $\tau = 1/2$, we see that the half-life of a shock to GDP is around 35 quarters, with little difference between the two. However, when there is a negative moving average component, $J_1^{.5}$ indicates a much longer half life than $J_0^{.5}$. On the other hand, the largest

[12] The data are taken from FRED. The web site address is http://www.stls.frb.org/fred.

autoregressive root is appropriate for evaluating these statistics for τ closer to 1. Hence when $\tau = .8$, both J_0^τ and J_1^τ suggest that it will take about 50 quarters for 80% of the effect of the shock to dissipate.

4. EMPIRICAL ANALYSIS OF REAL EXCHANGE RATES

One issue that arises frequently in the analysis of the exchange rate (real or nominal) is whether combining the low volatility data before the Bretton Woods agreement with data which are more volatile after the agreement will affect the size of unit root tests.[13] This issue of a break in variance was studied by Hamori, S. and A. Tokihisa (1997) for the normalized least squares estimator. The authors find that the break fraction and the relative variance in the two regimes will enter the limiting distribution of the test statistic, and in monte carlo experiments, combining the data of the two regimes will lead to over-rejections of the unit root tests. Although no formal analysis is available for other test statistics, there is little doubt that the qualitative conclusion will generalize. The focus of most exchange rate analysis on one regime is justified.

Table 5 presents estimates of ARMA(1,1) models for the nominal and the real exchange rate series as measured by the consumption deflator. The data are quarterly for the period 1973:1-1997:2, taken from the *OECD* sectoral database. Results using the *GDP* deflator are similar and not reported. For real exchange rates, the moving average component is, in most cases, estimated to be positive with t statistics larger than 1.6 in absolute value. Recall that Engel's basic premise was that Δq_t could have a large negative moving-average component. Evidence for this is found in only two countries, Australia and Korea. The U.K. data, which was the basis of Engel's analysis, clearly does not exhibit a negative θ. Since there is hardly any evidence of a negative θ in either q_t or Δq_t, size distortions of unit root tests should not be an issue. Applying the new (and old) tests, we can only reject a unit root in the real exchange rate for Canada, and only marginally. These results are reported in Table 6.

With 25 years of data, unit root tests may indeed have low power in providing a precise $I(1)/I(0)$ classification (see, e.g., Shiller, R.J. and P. Perron (1985), and Perron, P. (1991)). For this reason, we consider the measure of persistence J_0^τ along with bootstrapped standard errors.[14] With the exception of Korea, Greece, and Portugal, the half life of real exchange shocks is between nine and fifteen quarters, in line with the consensus estimate of 4.5 years from panel studies. Evaluating τ at .8 gives a clearer picture

[13] See, for example, the discussion in Froot, K. and K. Rogoff (1995), Section 2.3.5.

[14] Because of the lack of a negative moving-average component, estimates for J_1^τ are similar.

TABLE 5.

ARMA(1,1) Estimates of α and θ for Nominal and Real Exchange Rate: 1973:1-1997:2.

Country	s_t		q_t		Δq_t	
	α	θ	α	θ	α	θ
Australia	0.939	0.193	0.867	0.216	0.601*	-0.581*
Austria	0.892	0.380	0.901	0.354	0.605*	-0.444
Canada	0.949	0.426	0.951	0.425	-0.590*	0.919*
Denmark	0.943	0.363	0.913	0.375	-0.081	0.421
France	0.945	0.455	0.902	0.443	-0.086	0.490*
Germany	0.894	0.383	0.909	0.361	0.601*	-0.436
Greece	0.954	0.192	0.952	0.182	0.802	-0.735
Ireland	0.947	0.310	0.851	0.290	-0.248	0.466
Italy	0.939	0.412	0.890	0.423	-0.365*	0.702*
Japan	0.877	0.399	0.879	0.403	-0.009	0.368
Korea	0.968	0.364	0.935	0.263	0.719*	-0.466*
Luxembourg	0.939	0.409	0.929	0.384	-0.580	0.862
Netherlands	0.910	0.364	0.909	0.339	0.534	-0.314
Norway	0.897	0.364	0.870	0.365	-0.400	0.675*
Portugal	1.002	0.358	0.940	0.368	-0.318	0.619*
Sweden	0.920	0.378	0.924	0.295	0.514	-0.316
Switzerland	0.861	0.367	0.870	0.358	-0.680	0.915
U.K.	0.900	0.245	0.882	0.231	-0.184*	0.359*

For s_t and q_t, $\hat{\theta}$ is always significant at the two-tailed 10% level. For Δq_t, significant estimates are marked with an aseterik.

of the relative persistence across countries. Japan has the fastest speed of adjustment, with 80% completed in 15 quarters. This is followed by Canada, Ireland, Sweden, France and Italy. Adjustments in the remaining countries take over 20 quarters to complete, with Greece and Korea being the outliers.

5. CONCLUSION

From an econometric perspective, C. Engel's conclusion that long-run PPP may not hold is valid because we fail to reject a unit root in the real exchange rate. While the size issue being raised is a valid methodological problem, it appears not relevant to the exchange rate data we investigated. His result may be an artifact of the imprecise estimates of the parameters used to calibrate the model, leading to implied values of θ in the simulated data that are much closer to -1 than the observed data.

TABLE 6.

Statistics for the Exchange Rates.

Country	$\log(s_t)$					$\log(q_t)$				
	DF^{GLS}	MZ_α^{GLS}	$\hat{\alpha}$	$J_0^{.5}$	$J_0^{.8}$	DF^{GLS}	MZ_α^{GLS}	$\hat{\alpha}$	$J_0^{.5}$	$J_0^{.8}$
Australia	-1.334	-4.082	0.956	17	37	-1.981	-6.947	0.923	10 (3.0)	21 (7.0)
Austria	-2.147	-9.244	0.927	12	20	-2.035	-8.411	0.933	12 (3.4)	21 (6.9)
Canada	-2.068	-11.466	0.955	18	24	-2.305	-17.598	0.945	16 (3.2)	19 (4.6)
Denmark	-1.802	-6.581	0.955	18	31	-2.046	-8.232	0.940	14 (4.2)	23 (7.9)
France	-1.931	-6.919	0.955	18	29	-2.352	-10.342	0.928	12 (3.1)	18 (6.0)
Germany	-2.191	-10.298	0.928	12	20	-2.094	-8.532	0.935	13 (3.8)	22 (7.3)
Grece	-1.077	-2.589	0.972	26	58	-1.201	-3.234	0.965	21 (6.1)	46 (13.0)
Ireland	-1.593	-5.682	0.961	20	35	-2.199	-8.885	0.903	8 (2.2)	17 (5.1)
Italy	-1.609	-5.925	0.964	22	37	-2.383	-12.653	0.918	10 (2.7)	17 (5.3)
Japan	-2.511	-13.076	0.914	10	16	-2.578	-15.633	0.906	9 (2.2)	14 (4.3)
Korea	-1.490	-5.735	0.974	29	45	-1.119	-5.723	0.988	56 (15.2)	79 (20.7)
Luxembourg	-1.830	-7.018	0.953	17	29	-1.850	-7.039	0.948	16 (4.5)	27 (8.8)
Netherlands	-2.158	-10.083	0.931	12	20	-2.033	-8.843	0.938	13 (3.8)	23 (7.4)
Norway	-2.021	-7.995	0.936	13	23	-2.223	-9.384	0.922	11 (3.9)	19 (8.6)
Portugal	-0.872	-2.665	0.984	47	83	-1.843	-8.240	0.949	17 (5.0)	25 (10.0)
Sweden	-2.019	-8.485	0.940	14	24	-1.959	-8.082	0.939	14 (3.8)	24 (7.7)
Switzerland	-2.217	-10.585	0.924	11	19	-2.312	-10.605	0.919	10 (2.8)	18 (5.7)
U.K.	-1.679	-5.829	0.940	13	28	-1.991	-7.619	0.922	10 (2.8)	21 (6.5)

When a process has a unit root but has a negative moving average component, there are steps one can take to minimize size distortions while retaining power in unit root tests. The main ingredient is the use of a new information criterion, the $MAIC$, to select the autoregressive lag length. When a precise $I(1)/I(0)$ classification is not warranted, the use of the $MAIC$ still allows better estimates of measures of persistence related to the half-life of a shock. Our estimates put this half life of shocks to real exchange rate between nine and fifteen quarters, though there are more variations in the time required to complete 80% of the adjustments.

There remains the question of whether we should care if a non-stationary component in q_t exists? This issue is of independent interest because it is relevant whenever testing a variable which has subcomponents (such as the CPI and industrial production) is at stake. The answer depends on the objective of the exercise. Take industrial production. If an economist was asked "Are all sectors stationary?", then he should document as clearly as possible unit root tests results on all sectors. But if this economist was asked "is industrial production non-stationary", then there is no value-added in knowing if there is a permanent component in, say, the output for shoelaces. One might think otherwise if it was the production of automobiles rather than shoelaces that has a permanent component. But if the variations in automobiles are important enough, they will be reflected in industrial production anyway. In the end, unit root tests on the components are neither necessary nor sufficient for establishing a unit root in the aggregate variable. On the other hand, if we were interested in the source of the unit root in the aggregate, analysis of the components will be necessary. But establishing the existence and the source are two different questions.[15] In the case of the real exchange rate, little is lost from not knowing that a permanent component in y_t exists if all we want to know is whether the real exchange rate has a unit root.

APPENDIX A

Test Statistics

The DF^{GLS} test due to Dickey, D.A. and W.A. Fuller (1979), Said, S.E. and D.A. Dickey (1984) and Elliott, G., T.J. Rothenberg, and J.H. Stock (1996) is the t statistic on $\hat{\beta}_0$ from the augmented autoregression:

$$\Delta \tilde{z}_t = \beta_0 \tilde{z}_{t-1} + \sum_{j=1}^{k} \beta_j \Delta \tilde{z}_{t-j} + e_{tk}, \tag{A.1}$$

[15] Such an analysis was provided by Engel, C. (1999).

where $\tilde{z}_t = z_t - \hat{\beta}' d_t$, $\hat{\beta}$ is the GLS estimate of the coefficients on the deterministic terms d_t. That is, let $z_t^{\perp} = z_t - \bar{\alpha} z_{t-1}$ for $t = 2, ..., T$ and $z_1^{\perp} = z_1$, and let $d_t^{\perp}$ be similarly defined. Then $\hat{\beta}$ is obtained as the OLS estimate from a regression of $z_t^{\perp}$ on $d_t^{\perp}$. The non-centrality parameter is specified as $\bar{\alpha} = 1 + \bar{c}/T$ with $\bar{c} = -7.0$ when $d_t = \{1\}$ and $\bar{c} = -13.5$ when $d_t = \{1, t\}$.

The Phillips-Perron test is

$$Z_\alpha = T(\hat{\alpha} - 1) - (s^2 - s_v^2)(2T^{-2} \sum_{t=1}^{T} \ddot{z}_{t-1}^2)^{-1},$$

where $\ddot{z}_{t-1}$ are the residuals from an OLS regression of z_{t-1} on d_t, $t = 1, \ldots, T$, and $\hat{\alpha}$ is the least squares estimate from the regression

$$z_t = \beta d_t + \alpha z_{t-1} + v_t. \tag{A.2}$$

Also $s_v^2 = T^{-1} \sum_{t=1}^{T} \hat{v}_t^2$ and s^2 is a consistent (under the null hypothesis) estimate of the spectral density at frequency zero the first-differences of the stochastic component.

The Modified Phillips-Perron test MZ_α is

$$MZ_\alpha = \frac{T^{-1} \hat{z}_T^2 - s_{AR}^2}{2T^{-2} \sum_{t=1}^{T} \hat{z}_{t-1}^2} \approx Z_\alpha + \frac{T}{2}(\hat{\alpha} - 1)^2,$$

where $\hat{z}_t$ are the OLS residuals from a regression of z_t on d_t, $t = 0, \ldots, T$. The autoregressive estimate of the spectral density at frequency zero of v_t, is defined as:

$$s_{AR}^2 = s_{ek}^2 / (1 - \hat{\beta}(1))^2, \tag{A.3}$$

where $\hat{\beta}(1) = \sum_{i=1}^{k} \hat{\beta}_i$, $s_{ek}^2 = T^{-1} \sum_{t=k+1}^{T} \hat{e}_{tk}^2$, with $\hat{\beta}_i$ and $\{\hat{e}_{tk}\}$ obtained from the following autoregression estimated by OLS

$$\Delta z_t = \gamma' d_t + \beta_0 z_{t-1} + \sum_{j=1}^{k} \beta_j \Delta z_{t-j} + e_{tk}.$$

The GLS version, MZ_α^{GLS} is obtained using GLS detrended data, that is using $\tilde{z}_t$ instead of $\hat{z}_t$ and constructing s_{AR}^2 using the regression (A.1).

Engel's Model and Parameterizations

The model is:

$$\begin{aligned} \Delta y_t &= a u_t, \\ \Delta s_t &= -\delta(s_t + p_t^{*T} - p_t^T) + b u_t + c v_t, \\ \Delta(p_t^T - p_t^{*T}) &= \gamma(s_t + p_t^{*T} - p_t^T) + d\epsilon_t + f v_t + g u_t, \end{aligned}$$

where u_t, v_t, ϵ_t are *i.i.d.* $N(0,1)$ and are mutually uncorrelated. This implies for $x_t = s_t + p_t^{T*} - p_t^T$,

$$x_t = \phi x_{t-1} - d\epsilon_t + (c-f)v_t + (b-g)u_t,$$

with $\phi = 1 - \delta - \gamma$. Thus,

$$\begin{aligned} \sigma_w^2 &= a^2, \\ \sigma_m^2 &= (c-f)^2 + (b-g)^2 + d^2, \\ \sigma_{wm}^2 &= a(b-g) \end{aligned}$$

and θ determined according to (6).

Table A.1 Structural Parameters ($\times 100$) and Implied Values of ϕ, θ, and σ_w^2/σ_m^2.

Case	ϕ	θ	σ_w^2/σ_m^2	a	d	c	δ	b	g
1	0.919	-0.991	0.012	0.573	1.129	5.077	8.038	0.109	0.611
2	0.976	-0.999	0.000	0.100	1.890	4.150	2.400	0.109	0.611
3	0.952	-0.982	0.141	1.800	1.550	4.560	4.761	0.109	0.611
4	0.919	-0.925	3.750	10.00	1.129	5.077	8.038	0.109	0.611
5	0.850	-0.892	0.937	5.000	1.129	5.077	15.00	0.109	0.611
6	0.850	-0.971	0.037	1.000	1.129	5.077	15.00	0.109	0.611
7	0.999	-0.999	0.012	0.573	1.129	5.077	0.100	0.109	0.109
8	0.839	-0.982	0.012	0.573	1.129	5.077	8.038	0.109	0.109

In the above, $\gamma = 0$ except in case 8 when $\gamma = .08$. The parameter f is fixed at 0.000632. Cases 1,2 and 3 are from Tables 2 and 4 of Engel, C. (1999).

REFERENCES

Bergin, P. and R. Feenstra, 1999, Pricing to market, staggered contracts, and real exchange rate persistence, *NBER WP* 7026.

Blough, S., 1992, The relationship between the power and the level of generic unit root tests in finite samples. *Journal of Applied Econometrics* **7**, 295–308.

Campbell, J. Y. and P. Perron, 1991, Pitfalls and opportunities: What macroeconomists should know about unit roots. *NBER Macroeconomic Annual* **Vol. 6**, M.I.T. Press, pp. 141–201.

Clark, P. K., 1988, Nearly redundant parameters and measures of persistence in economic time series. *Journal of Economic Dynamics and Control* **12**, 447–461.

Cochrane, J. H., 1991, A critique of the application of unit root tests. *Journal of Economic Dynamics and Control* **15**, 275–284.

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

Elliott, G., T. J. Rothenberg, and J. H. Stock, 1996, Efficient tests for an autoregressive unit root. *Econometrica* **64**, 813–836.

Engel, C., 1999, Accounting for U.S. real exchange rate changes. *Journal of Political Economy* **107**, 507–538.

Engel, C., 1999, Long run PPP may not hold after all. *Journal of International Economics.*

Faust, J., 1996, Near observational equivalence and the theoretical size problems with unit root tests. *Econometric Theory* **12**, 724–731.

Froot, K. and K. Rogoff, 1995, Perspectives on PPP and long-run real exchange rates, *in* G. Grossman and K. Rogoff (eds), *Handbook of International Economics*, North Holland.

Hamori, S. and A. Tokihisa, 1997, Testing for a unit root in the presence of a variance shift. *Economic Letters* **57**, 245–253.

Mcgrattan, E., A. Chari, and P. Kehoe, 1998, Can a sticky price model generate volatile and persistent real exchange rates. Federal Reserve of Minneapolis Staff Report 223.

Nabeya, S. and P. Perron, 1994, Local asymptotic distribution related to the AR(1) model with dependent errors. *Journal of Econometrics* **62**, 229–264.

Ng, S. and P. Perron, 1995, Unit root tests in ARMA models with data dependent methods for the selection of the truncation Lag. *Journal of the American Statistical Association* **90**, 268–281.

Ng, S. and P. Perron, 1997, Estimation and inference in nearly unbalanced and nearly cointegrated systems. *Journal of Econometrics* **79**, 54–81.

Ng, S. and P. Perron, 2001, Lag length selection and the construction of unit root tests with good size and power. *Econometrica, forthcoming.*

Perron, P., 1991, Test consistency with varying sampling frequency. *Econometric Theory* **7**, 341–368.

Perron, P., 1996, The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors. *Journal of Econometrics* **70**, 317–350.

Perron, P. and S. Ng, 1996, Useful modifications to unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* **63**, 435–465.

Perron, P. and S. Ng, 1998, An autoregressive spectral density estimator at frequency zero for nonstationarity tests. *Econometric Theory* **14**, 560–603.

Phillips, P. C. B. and P. Perron, 1988, Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.

Said, S. E. and D. A. Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–607.

Schwert, G. W., 1989, Tests for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics* **7**, 147–160.

Shiller, R. J. and P. Perron, 1985, Testing the random walk hypothesis: power versus frequency of observation. *Economic Letters* **18**, 381–386.