NONSTATIONARITY AND LEVEL SHIFTS WITH
AN APPLICATION TO PURCHASING POWER PARITY

Pierre Perron
Princeton University and C.R.D.E.

Timothy J. Vogelsang
Princeton University

Econometric Research Program
Research Memorandum No. 359

Revised January 1991

ABSTRACT

This study considers testing for a unit root in a time series characterized by a structural change in its mean level. The analysis is in the spirit of Perron (1990a) who showed that the existence of such a shift in a stationary time series biases the usual tests for a unit root towards non–rejection. The approach is, however, different given that we suppose the date of the change to be unknown. The statistic of interest is then the minimal t–statistic over all possible break points in regressions similar to those proposed in Perron (1990a). We derive and tabulate the asymptotic distributions of interest. However, most of the emphasis is given to the tabulation of finite sample critical values using simulation experiments. Particular attention is given to the effect, on the finite sample critical values, of various procedures to select the appropriate order of the estimated autoregressions. We apply our procedure to analyze the issue of purchasing power parity between the US and the UK, and also between the US and Finland whose real exchange rates are characterized by apparent shifts in level when using particular price indices.

Key words :   Unit root, Hypothesis testing, Structural change, Simulation experiment, Stochastic trends, Deterministic trends.

Pierre Perron
Department of Economics
Princeton University
Princeton, NJ, 08544

Timothy J. Vogelsang
Department of Economics
Princeton University
Princeton, NJ, 08544

1. INTRODUCTION.

Testing for the presence of a unit root in a time series of data has become a topic of great concern to economists. Since the seminal work of Nelson and Plosser (1982), the view that most macroeconomic variables are best construed as exhibiting some kind of stochastic nonstationarity has become prevalent. The unit root hypothesis has far reaching implications with respect to both economic theory and the interpretation of empirical evidence. Indeed, under the unit root hypothesis random shocks have a permanent effect on the system, i.e. fluctuations are not transitory.

To illustrate these implications, consider the issue raised by the Purchasing Power Parity (PPP) hypothesis. A mild version of this hypothesis asserts that, given two countries, movements in the nominal exchange rate and movements in their respective price indices adjust over time such as to leave a constant relative purchasing power between the two countries (see, e.g., Dornbusch (1988)). In statistical terms, this hypothesis implies that fluctuations in the real exchange rate between any two given countries are stationary, i.e. that the shocks to this relation have temporary effects. On the other hand, a sufficient condition for a violation of PPP is that the real exchange rate is characterized by the presence of a unit root. In this case, shocks have permanent effects and there is no tendency to return to a stable value. This implication has been used by many authors to test the PPP hypothesis (e.g., Enders (1988, 1989), Corbae and Ouliaris (1988) and Taylor and McMahon (1988)). In this framework, failure to reject the null hypothesis of a unit root is taken as evidence against PPP.

In a recent study, Dornbusch and Vogelsang (1990) provide an empirical analysis of this issue using annual data over a long period. The analysis is performed with the real exchange rate between the US and many countries using (when possible) two price indices, a Consumer Price Index (CPI) and a GNP (or GDP) Deflator. Unlike many recent studies, the unit root hypothesis for the real exchange rate is usually rejected. This rejection can be explained by the higher power of the tests when a long span of data is used (see Perron (1990b)). However, in some instances the unit root is rejected with one price index but not the other.

To illustrate the issue, consider first Figure 1 which presents a graph of the real exchange rate between the US and the UK. Panel A is the real exchange rate based on the CPI, and Panel B is a similar series based on the GNP deflator. A standard Dickey–Fuller

(1979) test applied to the GNP deflator–based series rejects the unit root hypothesis at a high significance level (t–statistic of –3.23 with 8 lags of first–differences of the data added) [1]. However, the same test applied to the CPI–based measure implies a non–rejection (t–statistics of –1.59 with 4 additional lags). As can be seen from the graph, the CPI–based measure appears to exhibit a permanent change in level at some point between 1940 and 1950. Such is not the case with the GNP Deflator–based series. Figure 2 presents a similar example using the real exchange rate between the US and Finland [2]. Here, the opposite result holds. The unit root is easily rejected using the CPI–based measure (t–statistic of –5.74 with 1 lag) but not with the GDP–based measure (t–statistic of – 1.62 with 2 lags). As seen from the graph, the GDP based measure appears to exhibit a permanent decrease in level after 1940.

In the examples depicted above, the non–rejection of the unit root hypothesis is associated with an apparent permanent change in the level of the series. Such issues have recently been analyzed by Perron (1990a). In that paper it was shown that if a series is characterized by stationary fluctuations with a mean that exhibit a one–time permanent change in level, the usual tests for a unit root are biased towards non–rejection. Perron (1990a) also proposed a class of statistics to test the unit root hypothesis allowing for the possible presence of a one–time change in mean. The distribution theory underlying the derivation of the critical values assumed that the date of the change in mean was exogenous (or uncorrelated with the data). In the examples discussed above, this assumption appears inappropriate as it is difficult to single out any major exogenous event that could have caused the change in mean for one particular version of the real exchange rate measure while not for the other. It appears more plausible to view the possible change in mean as occurring at an unknown date [3]. The distribution theory underlying the unit root test in this case is, however, different.

The aim of this paper is to provide a class of statistics to test for the unit root hypothesis allowing for a possible change in the level of the series occurring at an unknown date. The models considered are similar to those in Perron (1990a), namely an "additive outlier model", appropriate when the change in level is sudden, and an "innovational outlier model", appropriate when the change in level occurs gradually. The class of statistics considered is similar to those in Banerjee, Lumsdaine and Stock (1990), Zivot and Andrews (1990) and Perron (1990c). They are based on the minimal value, over all possible break points, of the t–statistic on the sum of the autoregressive coefficients in the appropriate augmented autoregression . We also consider a class of statistics based on the

same t–statistic but now evaluated at the break point that yields the lowest (or highest) t–statistic on the coefficient associated with the change in mean. This permits imposing the mild a priori imposition of a one–sided change in mean and allows tests with possibly higher power.

Though we derive the asymptotic distribution of the statistics introduced, we put more emphasis on the finite sample distribution obtained using simulation experiments. In this context, we provide an extensive analysis of the effects of different procedures to select the order of the appropriate autoregression on the finite sample critical values. We consider the following procedures for selecting this truncation lag parameter : a) fixed , b) chosen according to a significance test on the last included lag (given a prespecified maximum) , c) chosen according to an F–test on additional lags (up to some prespecified maximum) , and d) chosen to minimize the relevant t–statistic for testing the unit root hypothesis. In each case, we present tabulated critical values for both the "additive outlier model" and the "innovational outlier model".

The plan of the paper is as follows. Section 2 discusses the class of models considered and the test statistics analyzed. The asymptotic distribution of the statistics under the null hypothesis of a unit root process is presented in Section 3. The finite sample critical values and the simulation experiments used to obtain them are discussed in Section 4. In Section 5, we apply our procedure to the two real exchange rate examples discussed previously. Section 6 provides a possible explanation and interpretation for the apparent change in mean in these series. Finally, Section 7 contains concluding comments and an appendix some mathematical derivations.

2. THE MODELS AND TEST STATISTICS.

We briefly discuss, in this section, the models and statistical procedures involved in testing for a unit root allowing for the possibility of a change in the level of the series. The reader is referred to Perron (1989, 1990a) for a more detailed discussion and interpretation. Throughout this paper, we assume that there is at most one change in the mean of the trend function. We denote the date of break, should it occur, by T_B with $1 < T_B < T$ where T is the sample size.

The first model is labelled, following the literature on time series with outliers (e.g. Tiao (1985)), the "additive outlier model" (AO). In this case, the change is assumed to take effect instantaneously. In particular, the effect of the change on the level of the series of interest $\{y_t\}$, say, does not depend on the dynamics exhibited by the correlation structure of $\{y_t\}$. Under the null hypothesis of a unit root, this model can be parameterized as :

$$y_t = \delta D(TB)_t + y_{t-1} + w_t \ , \qquad\qquad (t = 2, ..., T) \qquad\qquad (1)$$

where $D(TB)_t = 1$ if $t = T_B + 1$ and 0 otherwise. Here, and throughout this paper, $y_1 = y(1)$ is either a fixed constant or a random variable. The sequence of errors $\{w_t\}$ is, for simplicity, specified to be a stationary and invertible ARMA(p, q) process. More specifically, $A^*(L)w_t = B(L)e_t$ where e_t is i.i.d. $(0, \sigma^2)$ with $A^*(L)$ and $B(L)$ being polynomials in L of order p and q, respectively. The roots of $A^*(L)$ and $B(L)$ are assumed to be strictly outside the unit circle. Under model (1), the mean of the series is y(1) up to time T_B and $y(1) + \delta$ afterwards. Under the alternative hypothesis, the series $\{y_t\}$ does not contain a unit root and can be characterized by the following specification :

$$y_t = c + \delta DU_t + v_t \ , \qquad\qquad (t = 2, ..., T) \qquad\qquad (2)$$

where $DU_t = 0$ if $t \leq T_B$ and 1 otherwise. The sequence of errors $\{v_t\}$ is likewise a stationary and invertible ARMA(p + 1, q) process of the form $A(L)v_t = B(L)e_t$. The mean of the series is c up to time T_B and $c + \delta$ afterwards. The null hypothesis specified by

model (1) is a special case of the general specification (2) when c = y(1) and A(L) = (1 − L)A*(L). Hence, a natural testing strategy is to first remove an estimate of the deterministic part of the series (c + γDU_t) and to test whether the remaining noise is characterized by the presence of a unit root. For a fixed value of the break point T_B, this leads to the following two step procedure. First, remove the deterministic part of the series using the following regression estimated by OLS :

$$y_t = \mu + \delta DU_t + \tilde{y}_t \cdot \qquad\qquad (t = 1, ..., T) \qquad (3)$$

The test is then performed using the t−statistic for $\alpha = 1$ in the next regression (again estimated by OLS) :

$$\tilde{y}_t = \omega D(TB)_t + \alpha \tilde{y}_{t-1} + \Sigma_{i=1}^k c_i \Delta \tilde{y}_{t-i} + e_t \cdot \qquad (t = k + 2, ..., T) \qquad (4)$$

The reason for including the dummy variable $D(TB)_t$ in the second step regression will be discussed in more detail in Section 3. It is needed to ensure that the t−statistic on α in (4) has the same asymptotic distribution as in the "innovational outlier model". Note in particular that the asymptotic distribution is different if this regressor is omitted [4]. This testing strategy follows the method of Dickey and Fuller (1979) and Said and Dickey (1984) by approximating ARMA(p, q) processes by AR(k) processes. The method will be asymptotically valid provided k increases at a controlled rate as T increases (see Said and Dickey (1984)). We denote the t−statistic for testing $\alpha = 1$ in regression (4) as $t_{\hat{\alpha}}(AO, T_B, k)$ where T_B is the break date, k is the truncation lag parameter and AO indicates that we are applying the two step procedure appropriate for the "additive outlier model". The issues of interest are the specific procedures for choosing T_B and k. We return to this problem after a discussion of the "innovational outlier model".

Under the "innovational outlier model", the change is supposed to affect the level of the series $\{y_t\}$ gradually, i.e. there is a transition period. Though, in principle, the dynamic effect could take any form, a natural and simple way to model such a transition is to suppose that the economy responds to a "shock" to the trend function (here the change in mean) in the same way that it reacts to any other shocks. This assumption leads to the following specification under the null hypothesis of a unit root :

$$y_t = y_{t-1} + \psi(L)(e_t + \theta D(TB)_t) , \qquad (t = 2, ..., T) \qquad (5)$$

where $\psi(L) = A^*(L)^{-1}B(L)$ defines the moving average representation of the noise function with $A^*(L)$ and $B(L)$ as defined in (1). The immediate impact of the change in mean is θ while the long run impact is given by $\psi(1)\theta$. Under the alternative hypothesis of stationary fluctuations, the model is represented by :

$$y_t = a + \phi(L)(e_t + \delta DU_t) , \qquad (t = 2, ..., T) \qquad (6)$$

where $\phi(L) = A(L)^{-1}B(L)$ with $A(L)$ as defined in (2). In the stationary case, the immediate impact of the change in mean is δ, and the long run impact is $\delta\phi(1)$. Models (5) and (6) can be nested and approximated by the following finite order autoregressive model :

$$y_t = \mu + \delta DU_t + \theta D(TB)_t + \alpha y_{t-1} + \Sigma_{i=1}^{k} c_i \Delta y_{t-i} + e_t . \qquad (t = k + 2, ..., T) \qquad (7)$$

Again, this regression is in the spirit of Said and Dickey (1984) where an ARMA(p,q) model is approximated by a finite order AR(k) model. Under the null hypothesis of a unit root, α is equal to one (which also implies $\delta = 0$). Hence, the appropriate testing strategy is to use the t–statistic for testing $\alpha = 1$ when regression (7) is estimated by OLS. We denote such a t–statistic, obtained with given values of T_B and k, by $t_{\hat{\alpha}}(IO,T_B,k)$.

For both statistics $t_{\hat{\alpha}}(i,T_B,k)$ (i= AO, IO), the appropriate values of T_B and k are unknown. Hence, the testing procedure must take account of this fact. Consider first the modifications induced by the fact that T_B is unknown. As in Banerjee, Lumsdaine and Stock (1990), Zivot and Andrews (1990) and Perron (1990b), we analyze the behavior of the statistics $t_{\hat{\alpha}}(i,T_B^*,k) = \text{Inf}_{T_B \epsilon(k+2,T)} t_{\hat{\alpha}}(i,T_B,k)$ (i = AO, IO) where different specifications about the choice of k will be analyzed. This procedure is in the tradition of tests for structural change when the date of the change is assumed unknown. Note also that such a procedure is valid whether or not there is a change under the null hypothesis, i.e. we do not requite that T_B be identifiable (see, e.g., Davies (1987)).

We also investigate a second procedure where T_B is chosen such as to minimize the t–statistic for testing $\delta = 0$ in regression (3) for the AO model and in regression (7) for the IO model. Given the symmetric nature of the problem, this is also equivalent to analyzing the behavior of the statistic when T_B is chosen to maximize the t–statistic for testing $\delta = 0$ in each regression. We denote the resulting statistics by $t_{\hat{\alpha}}(i, T_B(\hat{\delta}), k)$ ($i = $ AO, IO). This procedure allows the imposition of the mild a priori restriction of a one–sided change in the mean level of the series while still treating the date of the break point as unknown. As we will see in later sections, this may allow a substantial increase in power relative to the other procedure.

For each of the two procedures for selecting T_B, we analyze different ways of selecting the truncation lag parameter k. The first method follows the theoretical development in Said and Dickey (1984) by treating k as a fixed function of the sample size T. Given that in finite samples such a rule does not restrict the choice of k, in practice this procedure simply specifies a choice of k that is independent of the data, i.e. determined a priori. We denote the statistic of interest under this procedure by $t_{\hat{\alpha}}(i, j, \bar{k})$ ($i = $ AO, IO ; $j = T_B^*, T_B(\hat{\delta})$).

In practice, the truncation lag parameter in an estimated autoregressions is rarely chosen as a fixed function of the sample size T, i.e. in such a way that the choice of k is uncorrelated with the data. Rather, it is usually chosen by some test statistic on the significance of the lagged first–differences of the data. We consider two such procedures. The first is that used in Perron (1989, 1990a). Here, for any given value of T_B, k is chosen such that the coefficient on the last included lag of the first–differences of the data is significant (at some significance level β) and the coefficient on the last included lag in higher order autoregressions are insignificant up to some a priori specified maximum order kmax, say. We use the fact that the asymptotic distribution of the t–statistic on these coefficients is standard normal to carry the inference. This procedure is justified in the pure AR case given that for an AR(p) process $c_i = 0$ for all $i > p$ and $c_i \neq 0$ for $i = p$. Hence, this procedure is implicitly estimating the true order of the autoregression assuming that an upper bound is known a priori [5]. We denote the resulting statistics by $t_{\hat{\alpha}}(i, j, k(t))$ ($i = $ AO, IO ; $j = T_B^*, T_B(\hat{\delta})$).

The other procedure investigated is similar and follows the empirical application in Said and Dickey (1984) where again an upper bound on the order is assumed to be known. This procedure uses an "F test" to assess the joint significance of the coefficients on the lagged first–differences of the data in the estimated autoregressions. The exact procedure is as follows. First, a maximum value of k, kmax, is specified. For a given value of T_B, the autoregression is estimated with kmax and (kmax – 1) lags. A $\beta\%$ one–tailed F–test is used to assess whether the coefficient on the kmaxth lag is significant, and if so, the value of k chosen is this maximum value. If not, the model is estimated with (kmax – 2) lags. The lag (kmax – 1) is deemed significant if either the F–test for (kmax – 2) versus (kmax – 1) lags or the F–test for (kmax – 2) versus kmax lags are judged significant based on the $\beta\%$ critical values of the χ^2 distribution. The procedure is repeated by lowering k until a rejection of the null hypothesis that additional lags are insignificant occurs or the lower bound k = 0 is attained. We denote the resulting statistics by $t_{\hat{\alpha}}(i,j,k(F))$ (i = AO, IO ; j = T_B^*, $T_B(\hat{\delta})$).

The last data dependent method investigated for selecting the order of the autoregression mimics the way T_B is chosen, namely to minimize the t–statistic for testing $\alpha = 1$ given some prespecified maximal value kmax. These statistics are formally defined as $t_{\hat{\alpha}}(i,j,k^*) = \inf_{k\epsilon(0,kmax)} t_{\hat{\alpha}}(i,j,k)$ (i = AO, IO ; j = T_B^*, $T_B(\hat{\delta})$). This is probably the procedure which uses the least information possible and presumably leads to tests with lowest power. It is useful, however, to consider such a case as it will clearly show that by choosing k to maximize the chance of rejection leads to critical values that are substantially higher (in absolute value). This suggests that inappropriate inference will be achieved (possibly even asymptotically) when using the critical values corresponding to the standard case [6]. Nevertheless, given that no simple asymptotic distribution appears manageable in this case and given the presumably low power associated with this procedure, we do not recommend it in practice.

3. THE LIMITING DISTRIBUTION OF THE STATISTICS.

Our strategy in discussing the asymptotic distribution of the statistics of interest is as follows. We first consider the case where under the null hypothesis the noise function is uncorrelated ; that is, in the "additive outlier model" (1) we specify $A^*(L) = B(L) = 1$. Similarly, in the "innovational outlier model" (5), $\psi(L) = 1$. In this case, both model yields the same data–generating process, namely :

$$y_t = \delta D(TB)_t + y_{t-1} + e_t \qquad y(1) = 0 \quad (t = 2, ..., T) , \qquad (8)$$

where $e_t \sim$ i.i.d. $(0, \sigma^2)$ and where the initial condition is set to 0 for simplicity. In the Appendix, we derive the limiting distribution of the statistics $t_{\hat{\alpha}}(i,j,k=0)$ (i = AO, IO ; j = T_B^*, $T_B(\hat{\delta})$) when no lags of the first–differences of the data are included in the autoregressions, i.e. $k = 0$ in (4) and (7). The asymptotic distributions derived under these conditions remain valid if additional correlation is present in the data–generating process when higher order autoregressions are applied. In the case where k is a fixed function of the sample size (such that $k \to \infty$ and $k^3/T \to 0$ as $T \to \infty$) we appeal to results of Said and Dickey (1984) to argue that the limiting distributions remain unchanged. The case where k is chosen according to a data based method is discussed below.

Consider first the asymptotic distribution of the statistics when T_B is chosen to minimize the t–statistic for $\alpha = 1$ in regressions (4) or (7). The first thing to note is that the same asymptotic distribution applies when considering the "additive" or "innovational" outlier model. Note, however, that this asymptotic equivalence holds only if the one–time dummy $D(TB)_t$ is included in the second step regression (4) in the case of the "additive outlier model". This is basically due to the fact that these models are similar except for some transitional effects which do not matter asymptotically.

Suppose first that k is a fixed function of the sample size and assume, as in Said and Dickey (1984), that $k^3/T \to 0$ as $T \to \infty$. The following result is proved in the Appendix using results in Perron (1989, 1990a) and Zivot and Andrews (1990). Let $W^*(\lambda,r)$ denote the projection residual of a standard Wiener process $W(r)$ (defined on the space $C(0,1)$) on the subspace generated by the functions $\{1, du(\lambda,r)\}$ where $du(\lambda,r) = 0$ if $r \leq \lambda$ and $du(\lambda,r) = 1$ if $r > \lambda$. We then obtain the following asymptotic representation under the null

hypothesis of a unit root process (either (1) or (5)) :

$$t_{\hat{\alpha}}(i, T_B^*, \bar{k}) \Rightarrow \inf_{\lambda \in \Lambda} \left\{ \int_0^1 W^*(\lambda, r) dW(r) \left[\int_0^1 W^*(\lambda, r)^2 dr \right]^{-1/2} \right\} \qquad (i = AO, IO) \quad (9)$$

where Λ is a closed subset of the interval (0, 1) and where $\Rightarrow$ denotes weak convergence in distribution. The fact that the space of values for λ is restricted to a closed subset of (0,1) means that, under this asymptotic interpretation, some trimming ought to be applied. Banerjee, Lumsdaine and Stock (1990) and Andrews (1990) suggest using a window $\lambda \in$ (.15, .85) which implies that, in finite samples, T_B can only take values in the range ([.15T], [.85T]) where [·] denotes the integer argument. This choice is arbitrary and, as in Zivot and Andrews (1990), we consider the largest "window" possible in both the theoretical derivations and the empirical applications, namely (k+2, T–1) [7]. We believe that this restriction on the interval specified by Λ can be relaxed and that we could consider the limiting distribution defined over the full interval [0,1]. Our proof, however, does not permit stating this explicitly as it relies on arguments in Zivot and Andrews (1990). These authors derive the limiting distribution of various components entering in the definition of the t–statistic and apply a continuous mapping argument to arrive at an expression similar to that of (9) in a related context. The application of this continuous mapping argument is valid if all the terms are bounded uniformly over λ. There is a need to trim the space of λ because some components are unbounded as λ approaches 0 or 1. However, the uniform boundedness of each component is only a sufficient condition to arrive at the desired result. In the case of the t–statistic discussed above, even though the individual components are unbounded as λ approaches 0 or 1, the t–statistic itself remains bounded. Hence, there seems to be no need to restrict the space Λ. To make this claim precise we would need to show tightness of the finite sample distribution as well as the convergence result presented in the appendix. Such a derivation is outside the scope of this paper. Note, however, that these comments do not apply to the limiting distribution of the statistics discussed below.

To tabulate the percentage points of the asymptotic distribution represented in (9) we proceeded by simulations in a manner similar to that of Perron (1989). First note that we can write (see Perron (1990a)) :

$$\int_0^1 W^*(\lambda, r) dW(r) = (1/2)(W(1)^2 - 1) - \lambda^{-1} W(\lambda) \int_0^\lambda W(r) dr$$

$$- (1 - \lambda)^{-1}(W(1) - W(\lambda)) \int_{\lambda}^{1} W(r) dr , \qquad (10)$$

and

$$\int_{0}^{1} W^*(\lambda,r)^2 dr = \int_{0}^{1} W(r)^2 dr - \lambda^{-1} (\int_{0}^{\lambda} W(r) dr)^2 - (1 - \lambda)^{-1} (\int_{\lambda}^{1} W(r) dr)^2. \qquad (11)$$

We used expressions (10) and (11) to obtain critical values via simulation methods. The procedure is briefly described as follows. First, we generate a sample of size 1,000 of i.i.d. $N(0,1)$ random deviates, $\{e_t\}$. We then construct sample moments of the data which converge weakly to the various functionals of the Wiener process involved in the representation of the asymptotic distribution. For example, as $T \to \infty$, $T^{-1/2} \Sigma_1^T e_t \Rightarrow W(1)$, $T^{-1/2} \Sigma_1^{T_B} e_t \Rightarrow W(\lambda)$, $T^{-3/2} \Sigma_1^T \Sigma_{j=1}^t e_t \Rightarrow \int_0^1 W(r) dr$, $T^{-1} \Sigma_1^T (\Sigma_{j=1}^{t-1} e_t) e_t \Rightarrow (1/2)(W(1)^2 - 1)$, etc. With a sample size of 1,000 and i.i.d $N(0,1)$ variates, we can expect the approximation to be quite accurate. Once the various functionals are evaluated, we construct the expressions in (10) and (11), and obtain one realization of the limiting distribution of the statistic $\int_0^1 W^*(\lambda,r) dW(r) [\int_0^1 W^*(\lambda,r)^2 dr]^{-1/2}$ for a given value of λ. For a given set of simulated errors $\{e_t\}$ we repeat this procedure for all values of $\lambda = j/1,000$; $j = 2, ..., 999$. The statistic of interest is then the minimum over all values of λ. The procedure is repeated 10,000 times and the critical values are obtained from the sorted array. The resulting critical values are presented in the rows labelled $T = \infty$ in Table 1 ("additive outlier model") and Table 2 ("innovational outlier model"). As mentioned before, the asymptotic distribution is the same in both cases.

Consider now the limiting distribution of the statistics $t_{\hat{\alpha}}(i,T_B^*,k(t))$ and $t_{\hat{\alpha}}(i,T_B^*,k(F))$ where the truncation lag parameter is chosen using a test of significance on the coefficients of the lagged first–differences of the data. The asymptotic result given in (9) remains valid if the noise function is that of an AR(p) process and the upper bound kmax is selected to be greater than p. This is basically due to the fact that, asymptotically, these procedures will select the appropriate order of the autoregressive process with probability one, and that the tests of significance on the coefficients of the lags are asymptotically independent of the t–statistic on α, see Hall (1990) for details. In the case where the noise function is a general ARMA (p,q) process no asymptotic results are

available. Our conjecture is that the asymptotic distribution given by (9) would remain valid using an argument similar to that of Said and Dickey (1984), namely by requiring the upper bound (kmax) to increase to infinity at a controlled rate as the sample size increases to infinity. The verification of such a conjecture is outside the scope of this paper.

Things are different when the truncation lag parameter k is chosen to minimize the value of the t–statistic for testing $\alpha = 1$. Here, no asymptotic results are yet available. The problems lies in the fact that the procedure for selecting k remains correlated with the t–statistic on α even asymptotically. There is also a presumption that the asymptotic distribution of $t_{\hat{\alpha}}(i, T_B^*, k^*)$ (i = AO, IO) remains affected by nuisance parameters (i.e. the particular correlation structure of the data) even asymptotically. As mentioned above, we present results concerning this statistic mainly to enquire about the extent to which data mining can create problems of inference, and such statistics are not recommended in practice.

Consider now the asymptotic distribution of the statistics $t_{\hat{\alpha}}(i, T_B(\hat{\delta}), \bar{k})$ (i = AO, IO) where T_B is now chosen to minimize the t–statistic for testing $\delta = 0$, the coefficient on the change in mean in regressions (3) or (7). Unlike the case where T_B is chosen to minimize $t_{\hat{\alpha}}$, the asymptotic distributions of the statistics in the "additive outlier model" and the "innovational outlier model" are different. We first consider the case of the "innovational outlier model". Again, the asymptotic distribution is the same when considering the behavior of the t–statistic in (7) with k = 0 under the model (8) as it is when k increases at a suitable rate as T increases and the general model (5) is considered. Derivations similar to those involved in the proof of Theorem 2 of Perron (1989) yields the following representation, valid whether $\delta = 0$ or not in (8) (see the Appendix for detail) :

$$t_{\hat{\alpha}}(IO, T_B(\hat{\delta}), \bar{k}) \Rightarrow \int_0^1 W^*(\lambda^*, r) dW(r) \left\{ \int_0^1 W^*(\lambda^*, r)^2 dr \right\}^{-1/2} , \qquad (12)$$

where $\lambda^* = \operatorname{argmin}_{\lambda \in \Lambda} Z(\lambda)$ with $Z(\lambda) = H(\lambda)/K(\lambda)^{1/2}$,

and $\quad H(\lambda) = A[\lambda W(1) - W(\lambda)] - B(\lambda)[(1/2)(W(1)^2 - 1) - W(1) \int_0^1 W(r) dr]$,

$$K(\lambda) = A[\lambda(1 - \lambda)A - B(\lambda)^2] ;$$

with $A = \int_0^1 W(r)^2 dr - (\int_0^1 W(r)dr)^2$ and $B(\lambda) = \lambda \int_0^1 W(r)dr - \int_0^\lambda W(r)dr$.

As before Λ denotes a closed subset of the interval $(0,1)$. To simulate the asymptotic distribution represented by (12) we proceeded as in the case of the tabulation of the asymptotic distribution described by (9). These critical values are presented in Table 4 in the row labelled $T = \infty$. Note that the same comments apply as before concerning the asymptotic distribution of the statistics when the truncation lag parameter is chosen according to a significance test on the coefficients of the lagged first–differences or is chosen to minimize the t–statistic for testing $\alpha = 1$.

Consider now the asymptotic distribution of $t_{\hat{\alpha}}(AO,T_B(\hat{\delta}),k)$ for the "additive outlier model", where again T_B is chosen to minimize the t–statistic for testing $\delta = 0$ in (3), denoted $t_{\hat{\delta}}$. The first thing to note is that $t_{\hat{\delta}}$ is unbounded as $T \to \infty$. However, a non–degenerate asymptotic distribution is obtained if we consider the scaled version $T^{-1/2} t_{\hat{\delta}}$. Hence, in this asymptotic framework, the procedure is valid if one considers choosing T_B such that the value of $T^{-1/2} t_{\hat{\delta}}$ is minimized. Of course, in finite samples, it makes no difference whether one minimizes $t_{\hat{\delta}}$ or $T^{-1/2} t_{\hat{\delta}}$. It is shown in the Appendix that (whether or not $\delta = 0$ in (8)) :

$$t_{\hat{\alpha}}(AO,T_B(\hat{\delta}),\bar{k}) \Rightarrow \int_0^1 W^*(\lambda^*,r)dW(r)\left\{\int_0^1 W^*(\lambda^*,r)^2 dr\right\}^{-1/2} , \tag{13}$$

where $\lambda^* = \text{argmin}_{\lambda \in \Lambda} Q(\lambda)$ with $Q(\lambda) = B(\lambda)/[K(\lambda)/A]^{1/2}$,

where A, $B(\lambda)$ and $K(\lambda)$ are as defined in (12). To tabulate the critical values of the asymptotic distribution we proceeded in a manner analogous to the previous cases. These critical values are presented in Table 3 in the row labelled $T = \infty$. Again, similar comments as expressed above apply concerning the asymptotic distribution of the statistic when the truncation lag parameter is chosen according to a significance test on the coefficients of the lagged first–differences or is chosen to minimize the t–statistic for testing $\alpha = 1$.

4. SIMULATIONS OF THE FINITE SAMPLE CRITICAL VALUES.

In this section, we present an extensive study of the finite sample distribution of the statistics $t_{\hat{\alpha}}(i,j,k)$ (i = AO, IO, ; $j = T_B^*$, $T_B(\hat{\delta})$) under the null hypothesis of a unit root using various procedures for selecting the truncation lag parameter k. We consider the following data generating process for the simulations :

$$y_t = y_{t-1} + e_t \; , \qquad\qquad y_1 = 0 \; . \qquad\qquad (14)$$

We set $y_1 = 0$ without loss of much generality (the effect of the initial condition being minor). Similarly, we impose $\delta = 0$ in (1) and $\theta = 0$ in (5) given that the statistics are asymptotically invariant to these parameters. There may be an effect in small samples, but it is likely to be small. Furthermore, the case of a null hypothesis specified with no structural change is the leading case of interest given that one may often wish to discriminate between a constant mean unit root process and a stationary process with a changing mean. We specify the sequence of innovations $\{e_t\}$ to be i.i.d. N(0,1) even though the finite sample distributions of the statistics are not invariant to the correlation structure of the data. The basic justification is that the asymptotic distributions are invariant to additional correlation in the data under some regularity conditions (except perhaps when k is chosen to minimize $t_{\hat{\alpha}}$). Hence, there is no loss in generality in specifying i.i.d. errors under this asymptotic interpretation. The implicit assumption is that, in finite samples, the introduction of additional lags of first–differences of the data would mostly eliminate the dependency of the distributions on the correlation structure. More importantly, by specifying i.i.d. errors we are better able to assess the relative impact on the finite sample distributions of various factors such as the length of the autoregression (if fixed), the finite sample correlation between the statistics to choose k and the t–statistic on α (in the $k(t)$ and $k(F)$ methods), and the effect of choosing k based on the value of the t–statistic on α. Using i.i.d. errors allows us to isolate the relative effects of each of these factors.

For all statistics considered, we present results for three sample sizes, namely T = 50, 100 and 150. When choosing k using a data dependent method, we specify the upper bound on the autoregression, kmax, to be 5 [8]. In the case where k is a fixed value, we present results for $k = 0, 2,$ and 5 . This allows us to analyze the effect of overparameterization on the finite sample critical values. When k is chosen according to a test on the coefficients of

the lagged first–differences, the size of the test is set at 10% [9] . Each set of results was obtained using 2,000 replications. The program was coded using the C language, and N(0, 1) random deviates were obtained from the routine RAN1 of Press et al. (1986). To minimize sampling variability, we used the same set of generated data for all cases that share a common value of T, the sample size.

Table 1 presents the results concerning the statistic $t_{\hat{\alpha}}(AO, T_B^*, k)$, i.e. the t–statistic for testing $\alpha = 1$ in the "additive outlier" regression (4) where the break point T_B is chosen to minimize this t–statistic. Consider first the asymptotic distribution. The first thing to note is that it is considerably shifted to the left compared to the case where the break point is assumed known. Consider for instance the 5th percentage point. From Perron (1990a), the asymptotic critical value when the break point is assumed to occur at mid–sample is –3.04. In contrast, when the break point is assumed unknown, Table 1 shows the corresponding critical value to be –4.44. Secondly, the asymptotic distribution is a good approximation to the exact distribution when k is fixed at 0 (though slightly less than in the "innovational outlier model" discussed below).

When the autoregression is overparameterized (k is greater than 0) the critical values increase substantially, especially in the left tail of the distribution. The extent of the increase is more pronounced for a small sample size (e.g., T = 50) than for a larger sample size (e.g., T = 150). This feature is interesting. Indeed, it is generally believed that test statistics of the Dickey–Fuller (1979) type have power that decreases as the order of the autoregression increases because more parameters need to be estimated and because of the loss of observations due to the need for additional initial conditions. The fact that the finite sample critical values increases as k increases implies that these effects may be counteracted to some extent when the increase in k results in overparameterization [10].

When k is chosen using a test of significance on the coefficients of the lagged first–differences (rows k = k(F) and k = k(t)), the critical values are substantially smaller than those corresponding to both the fixed k case and the asymptotic distribution, especially if the sample size is small. This is due to the correlation, in finite samples, between the statistics on the coefficients of the lagged first–differences and the t–statistic for testing $\alpha = 1$. As discussed above, this correlation disappears asymptotically. Our simulation results shows this correlation to vanish slowly. Hence, for common sample sizes, care must be taken in using the asymptotic critical values in this case. Also of interest is

the fact that the critical values using an F–test (row k = k(F)) to assess the joint significance of the coefficients on the lagged first–differences are somewhat higher (in the left tail of the distribution) than those corresponding to the case where a t–test on the coefficient of the last lag is used (row k = k(t)). The difference, however, becomes minor, when T = 150. This implies that, for small sample sizes, the former procedure may be more powerful. Finally, when k is chosen such that the t–statistic for testing $\alpha = 1$ is minimized (row k = k*), the critical values are substantially smaller than those in all other cases even for large sample sizes. This illustrates the extent to which data mining can affect inference. Indeed, selecting k using such a procedure and not properly adjusting the critical values would lead to tests with inflated sizes. However, once the critical values are properly chosen, such a procedure is likely to lead to tests with relatively low power and, hence, is not recommended.

Table 2 presents the critical values corresponding to the statistic $t_{\hat{\alpha}}(IO, T_B^*, k)$, i.e the t–statistic for testing $\alpha = 1$ using the "innovational outlier" regression (7) and T_B chosen to minimize this t–statistic. Many of the qualitative results are similar to the case of the "additive outlier model" presented in Table 1. Hence, we discuss only the main features and differences. First, the asymptotic distribution is very close to the finite sample distribution with k fixed at 0 (especially in the left tail of the distribution), even for a sample size as small as 50. Secondly, unlike in the "additive outlier model", the critical values with k fixed are relatively stable across various values of k. Third, even for small sample sizes there are no significant differences between the critical values when using either test of significance for choosing k (the k(t) or k(F) procedures). In general, the critical values are quite similar to those in the "additive outlier" case. Again, the method whereby k is chosen to minimize the t–statistic on α yields the smallest critical values.

We now turn to the results concerning the cases where the time of break T_B is chosen to minimize the t–statistic on the coefficient associated with the change in mean. Table 3 presents the results corresponding to the "additive outlier model" where T_B is chosen to minimize the t–statistic for testing $\delta = 0$ in regression (3). Recall that this procedure allows the a priori imposition of a one–sided change in the mean of the series (the results also apply if one were to maximize the t–statistic for testing $\delta = 0$). Hence, as stated above, we can expect possibly higher power through higher critical value (i.e. less negative). As shown in Table 3, this is indeed the case. For instance, the asymptotic 5%

critical value is −3.61 compared to −4.44 (and the 10% critical value is −3.27 compared to −4.19). Similarly, the finite sample critical values are uniformly higher (less negative) than those for the "additive outlier model" where T_B is chosen to minimize the t−statistic on α (see Table 1). Indeed much of the same qualitative comparisons across methods hold : the critical values decrease noticeably as k increases when the sample size is small (the differences are minor when T is large), the critical values from the procedure using the F−test on the coefficients of the lags are larger than those using the t−test when T = 50 but not significantly different when T is 100 or 150, and the procedure whereby k is chosen to minimize the t−statistic on α leads to the smallest (most negative) critical values.

Finally, Table 4 presents the simulation results concerning the statistics $t_{\hat{\alpha}}(IO, T_B(\hat{\delta}), k)$ using the "innovational outlier model" where T_B is chosen to minimize the t−statistic for testing $\delta = 0$ in (7). A feature of substantial interest is that the asymptotic critical values in the "innovational outlier model" are substantially smaller than in the "additive outlier model". For instance, the 5% critical value is −4.19 compared to −3.61. One would therefore expect higher power under the "additive outlier model". The other differences between the critical values in Tables 3 and 4 are similar to the differences discussed when comparing the results in Tables 1 and 2. In particular, the asymptotic distribution is, in general, a good approximation to the finite sample distribution when k is fixed and the critical values are not much influenced by the value of k when it is fixed. There are very little differences between the critical values using the F−test or the t−test on the coefficients of the lagged first−differences, and the critical values are smaller throughout (more negative) in the "innovational outlier" case than they are in the "additive outlier" case (though here the difference is more important).

The results presented in this section permit testing the unit root hypothesis allowing for a possible change in level using a wide variety of procedures. Whether the "additive" or "innovational" outlier model is appropriate depends, of course, on the particular series. Concerning the procedure to be adopted for each of these models, we recommend using a data dependent method for selecting the truncation lag parameter k whereby this lag is chosen according to a significance test on the coefficients of the lagged first−differences. Using a fixed k procedure is bound to produce test that may either have the wrong size (due to possible underspecification) or to lack power if the order is wrongly selected. Use of the data dependent procedure whereby k is chosen to minimize the t−statistic for testing α = 1 may provide a rough guide, but is open to difficulty of interpretation.

5. APPLICATION TO THE PURCHASING POWER PARITY HYPOTHESIS.

In this Section, we apply the proposed testing procedures to the two real exchange rates discussed in the introduction, namely the US/UK series based on the CPI and the US/Finland series based on the GNP deflator. The motivation for using a unit root test on the real exchange rate series as a test of the Purchasing Power Parity (PPP) hypothesis can be described as follows. First, consider the multiplicative model of the real exchange rate given by $E_t P_t^*/P_t = bw_t$, where E_t is the dollar price of the foreign currency, P_t is the U.S. price level, P_t^* is the foreign price level, b is a constant and w_t is a positive disturbance term with mean 1. Given that PPP is rarely considered a theory of exchange rate determination, we can view E_t, P_t and P_t^* as endogenously determined. Defining $E_t P_t^*/P_t$ to be the real exchange rate, R_t, we have $R_t = bw_t$, and taking the logarithm we obtain the linear specification $r_t = r + u_t$, where r_t is the logarithm of the real exchange rate, r is a constant and u_t is a mean zero disturbance term. The absolute version of PPP states that the real exchange rate is equal to one at every point in time (which would imply $r = 0$ and $u_t = 0$). Practical considerations imply a less strict interpretation. First, since only price indices are observed and not price levels, we can only test a relative version of PPP which states that r must be constant but imposes no restrictions on its level. Second, given the world's inherent stochastic environment, we would expect deviations from PPP over time. If we allow only transitory shocks, u_t is then a stationary stochastic process. It is this empirical implication of PPP which is of interest. Indeed, a situation in which shocks have permanent effects is in contradiction to the version of PPP just described. It is therefore of interest to test whether shocks to the real exchange rate exhibit a long term effect.

In practice, one must specify a general class of models in which these issues can be analyzed. For our purpose, it is useful to consider the process u_t as a member of the class of finite–order ARMA(p,q) models, i.e. $A(L)u_t = B(L)e_t$ where e_t is a white noise disturbance term. In this framework, we can test the null hypothesis that PPP does not hold by testing whether the autoregressive polynomial $A(L)$ contains a root on the unit circle. Given that the interest in testing PPP centers around the stochastic behavior of the noise function $\{u_t\}$, we may wish to specify a more flexible structure about the level of the real exchange

rate series. In particular, as Figures 1 and 2 suggest, one may want to allow for the possibility of a structural change in mean at some unknown date. Reasons for such a possible change will be investigated later. Using the methods described in this paper, one can test whether the noise function $\{u_t\}$ is characterized by the presence of an autoregressive unit root.

We applied the procedures proposed in earlier sections to the US/UK CPI based real exchange rate series and the US/Finland real exchange rate series based on the GNP deflator. Consider first applying the "additive outlier model" with the break point T_B chosen to minimize the value of the t–statistic on the sum of the autoregressive coefficients, i.e. $t_{\hat{\alpha}}(AO, T_B^*, k)$. The results are presented in Table 5 for three methods of choosing T_B and k. When all methods yield the same values, only one regression is reported, otherwise separate regressions are listed. The statistics of most interest are the estimates of α and their t–statistics as well as four sets of p–values in the last four columns (reported to the nearest 1%). The first set of p–values is obtained using the distribution of the t–statistic in the fixed k scenario using the empirically selected value of k. This amounts, basically, to reporting the p–values based on the asymptotic distribution corrected, to a first approximation, for small sample biases (except perhaps when using the procedure where k is chosen to minimize the t–statistic on α). The second set of p–values is obtained using the distribution of the t–statistic when both T_B and k are chosen to minimize the t–statistic on α (k = k* in Table 1). The third set of p–values correspond to the critical values of the t–statistic when k and T_B are chosen so that the last lag is significant according to a 10% two–sided t–test on its coefficient and the last lag in higher order autoregressions has insignificant coefficients (k = k(t) in Table 1). The last set of p–values are those corresponding to the case where T_B and k are chosen according to a 10% joint significance test on the coefficients of the lagged first–differences of the data (k = k(F) in Table 1). In all cases, we selected kmax = 5 and the p–values are obtained using the simulations performed for T = 100. As the results in Table 5 show, the unit root hypothesis can easily be rejected at the 5% level using any procedure for selecting k and T_B. Indeed, for both the US/Finland and US/UK series, all procedures yields the same values, namely $T_B = 1938$ and k = 1 for US/Finland and $T_B = 1943$ and k = 1 for US/UK [11].

Table 6 presents results applying the "innovational outlier model" and selecting T_B and k to minimize the t–statistic for testing $\alpha = 1$, i.e. using the statistic $t_{\hat{\alpha}}(IO,T_B^*,k)$. The presentation of the results follows that in Table 1. For the US/UK series, $T_B = 1944$ and k = 1 irrespective of the method used to select T_B and k, and the results again allow a rejection of the unit root at the 5% level. For the US/Finland series, $T_B = 1945$ (and k = 1) when using a test of significance of the coefficients of the lags and the p–value for the null hypothesis of a unit root is .07. When using no such test of significance, $T_B = 1938$ (k = 1) and the p–value is .03. Hence, in both cases the unit root is rejected. Since the noise function is stationary one can use the fact that the asymptotic distribution of $t_{\hat{\delta}}$ is standard normal and conclude that the US/Finland series is characterized by a significant decrease in mean, while the US/UK series is characterized by a significant increase in mean.

Tables 7 and 8 present the results when T_B and k are selected to minimize (in the case of US/Finland) or maximize (in the case of US/UK) the t–statistic on the coefficient associated with the change in mean, $t_{\hat{\delta}}$. The application of these procedures imply the a priori imposition of a one–sided change in mean occurring at an unknown date. Table 7 considers the "additive outlier model", i.e using the statistic $t_{\hat{\alpha}}(AO,T_B(\hat{\delta}),k)$. The same procedures are used to select k and T_B. For the US/Finland series the unit root is rejected at less than the 1% level using any procedure. This shows how the mild a priori imposition of a one–sided change can increase the power of the test substantially. The results are similar for the US/UK series with a rejection at close to the 1% level. Here the break point is estimated to be 1940 for the US/Finland series and 1946 for the US/UK series.

Table 8 presents similar results when using the "innovational outlier model". For the US/UK series, $T_B = 1944$ and k = 1 irrespective of the method used, and the unit root hypothesis is again easily rejected. For the US/Finland series, $T_B = 1945$ if a test of significance on the coefficients of lagged first–differences is used. Here again, the unit root can be rejected at the 5% level.

The results obtained strongly suggest that both the US/Finland real exchange rate based on the CPI and the US/UK real exchange rate based on the GNP Deflator are stationary series if allowance is made for the possibility of a one–time change in the mean of the series. This is contrary to the evidence obtained using a standard Dickey–Fuller test in which case the unit root cannot be rejected. Our results show the non–rejection to be due to the fact that no allowance was made for a change in mean.

6. PRELIMINARY EXPLANATION OF SHIFTS IN REAL EXCHANGE RATES.

In this Section we present a possible explanation for the shift in mean apparent in the real exchange rate series discussed in Section 5. For the sake of conciseness we focus on the US/Finland series. A similar argument could possibly explain the case of the US/UK real exchange rate.

The issue at hand is the apparent shift in the level of the US/Finland real exchange rate when GDP deflators are used as the price indices. When CPI's are used as price indices, the level of the US/Finland real exchange rate remains fairly stable (see Figure 2). Because there is no shift when the CPI's are used, we can deduce that the shift in the GDP deflator–based series is not due to changes in the nominal exchange rate. In fact, the nominal exchange rate was moving in opposite direction to the real exchange rate at the time of the shift. Thus the source of the shift must be from one of the GDP deflator series.

To determine which of the two deflator series, US or Finland, is the source of the shift we constructed GDP deflator–based real exchange rates between the US, Finland and the following countries: Canada, Japan, Sweden and the UK [12]. In the case of the US, all of the series, except for Japan, were stable and no shift in mean was present. However, in the case of Finland, all of the series, except for Japan, exhibited a shift in mean near the time of World War II [13]. We view this as evidence that the shift in mean was a phenomenon between Finland and the rest of the world. To further corroborate the fact that the Finland GNP deflator was the source of the shift and not the Finnish nominal exchange rate, we constructed CPI–based real exchange rates between Finland and the following countries: Canada, France, Italy and Sweden [14]. In all cases the level of the real exchange rate was stable with no apparent shift in mean [15].

As a preliminary attempt at explaining the behavior of Finland's real exchange rate, we discuss the issue in the context of a simple Ricardian free trade model taken from Jones (1979). In this model there are two countries which we refer to as Finland and the rest of the world (ROW). There are three goods produced in the world economy using a single input, labor. Good 2 is produced by both Finland and the ROW and is the numeraire. We assume that Finland is the sole producer of good 1 in which it has a comparative technological advantage. Likewise, the ROW is the sole producer of good 3. Production is described in terms of unit labor costs a_{Li} where i refers to the good. We assume that all

goods are traded and consumed by both Finland and the ROW, and that consumption varies little between Finland and the ROW.

Consider technical regress in good 1. At constant wages, the percentage change in the price of good 1 is positive and equal to the percentage change in the unit labor cost associated with that good, a_{L1}. With the price of good 1 higher the GDP deflator in Finland will be higher. However, since good 2 is the numeraire, the price of good 3 remains constant. Hence, the GDP deflator in the ROW will be unchanged. Therefore, the real exchange rate in Finland, measured in terms of domestically produced goods, will appreciate. On the other hand, due to the fact that Finland and the ROW both consume good 1, the CPI in both countries will rise leaving the real exchange rate, measured in terms of consumption, unchanged. So, with technical regress in good 1 we would expect to see a shift in the GDP deflator–based real exchange rate, but no change in the CPI–based real exchange rate.

Within the model just described, we suggest the following story. In the 1940's Finland's largest single export good was lumber and lumber by–products comprising 80–85% of export value (c.f. Bank of Finland (1947)). If we concentrate on other European countries, we could think, to a first approximation, of lumber as being specific to Finland. For our story to apply, there would have to be technical regress in Finland's lumber industry in the 1940's. A possibility is that Finland switched to a type of forest harvesting which takes into account the costs of depleting a scarce, but only slowly renewable resource such as forests. In this case we would see a reduction in the productivity of lumber related activities, i.e. technical regress. Of course such an explanation is preliminary in nature and no doubt a more extensive analysis is needed to fully account for the phenomenon described. Nevertheless, the point made is sufficient for our purpose in that it shows how it is possible to have a significant one time change in structure which can cause a shift in mean in some measured real exchange rate series. Such a change has a separate source from the regular temporal fluctuations in the real exchange rate series. It is these regular fluctuations that are of interest when testing issues such as purchasing power parity. Hence, this illustrates the need to separate this one–time change from the noise function as was done in the last section.

7. CONCLUDING COMMENTS.

This paper has presented a class of procedures to test the null hypothesis of a unit root in a time series of data that is possibly affected by a one–time change in its mean level. We considered two classes of models, labelled the "additive outlier model" and the "innovational outlier model". The former is best suited for series exhibiting a sudden change in mean while the latter is best suited if the change takes place gradually. The appropriateness of these models henceforth depends on the actual series being used. Nevertheless, our empirical applications showed that the results can be robust to alternative specifications.

For each of these models, we considered two type of statistics. The first considers the minimal value of the t–statistic on the sum of the autoregressive coefficients over all possible break points. This, in effect, allows testing the unit root hypothesis without any a priori assumption about either the location of the break point or the sign of the change in mean. In an attempt to provide statistical tests with higher power, we also analyzed a procedure whereby the break point is chosen to minimize (or maximize) the t–statistic on the coefficient of the change in mean, henceforth imposing the mild a priori assumption that the sign of the possible change is known (while its location remains unknown). This class of statistics allows an interesting sensitivity check on the obtained inference that may be a compromise between the more powerful case where the break point is assumed known and the least powerful case where no structure is imposed.

Finally, for each procedure suggested, we analyzed the effects on the critical values of different assumptions about the method used for choosing the truncation lag parameter in the estimated autoregressions. In practice, we recommend using a test of significance on the coefficients of the lagged first–differences. In particular, the k(t) procedure, whereby the order is selected by a test of significance of the last included lag, is quite easy to implement using standard regression output. Care, however, must be taken to correct for finite sample biases introduced by the correlation between the final statistics of interest and the t–tests on these additional lags. Our tabulated critical values should prove to be useful in this respect.

FOOTNOTES

[1] In this and other Dickey–Fuller tests reported in this Section, the t–statistics are constructed from a regression with a constant but no trend. Lags of first–differences are added to account for serial correlation. The number of such lags included was determined using a 10% two–sided significant test on the last included lag in the autoregressions (see Section 2 for more detail). The 5% and 1% critical values are –2.86 and –3.43, respectively.

[2] The US/UK real exchange rate series based on the CPI index is annual from 1892 to 1988. The series based on the GNP Deflator is also annual from 1869 to 1987. The source of the data is as follows. Nominal exchange rate series : 1869–1912 from Mitchell (1981), 1913–1949 from Board of Governors of the Federal Reserve (1943, 1976), 1950–1988 from International Monetary Fund (1988) ; Consumer Price Index : 1892–1969 from Mitchell (1981), 1970–1988 from International Monetary Fund (1988); GNP Deflator : 1868–1975 from Friedman and Schwartz (1982), 1976–1987 from International Monetary Fund (1988). The US/Finland real exchange rate series based on the CPI index is annual from 1900 to 1988, while that based on the GDP deflator is from 1900 to 1987. The source of the data is as follows. Nominal exchange rate series : 1900–1988 from the Bank of Finland ; Consumer Price Index : 1900–1985 from the Bank of Finland, 1986–1988 from the International Monetary Fund (1988) ; GDP Deflator : 1900–1985 from the Bank of Finland, 1986–1987 from International Monetary Fund (1988). The sources of the data for the US are as follows. GNP deflator : 1869–1975 from Friedman and Schwartz (1982), 1976–1988 from IMF (1988) ; CPI : 1860–1970 from the U.S. Bureau of Census (1976), 1971–1988 from IMF (1988).

[3] There is also some validity to the argument that the break point ought to be viewed always as an unknown parameter irrespective of the fact that one may be able to identify an exogenous event that is responsible for the change at a particular date (see, e.g. Christiano (1988)). However, the distribution theory obtained assuming a known break point may often be a good approximation to the appropriate distribution if no attempts were made to select the break point by systematically analyzing the data prior to this choice. See the discussion in Perron (1990c).

[4] Note that the second step regression (4) is different from the one proposed in Perron (1990a) where the dummy regressor $D(TB)_t$ was not included. The treatment of the

asymptotic distribution in the "additive outlier model" is, in fact, erroneous in Perron (1990a). However, all the results about the asymptotic distribution remain valid if this additional regressor is included. See section 3 and the appendix for more detail.

[5]   See Section 3 for a discussion of the validity of such a procedure, in an asymptotic framework, when the data is generated by an ARMA(p,q) process.

[6]   The same comments apply to the usual Dickey–Fuller (1979) test. If the truncation lag parameter k is chosen a priori or by a test of significance on the coefficients of the lagged first–differences, the t–statistic will have the asymptotic distribution tabulated in Fuller (1976). This will occur provided the test related to the choice of k is applied starting from a general overparameterized model successively eliminating insignificant lags (see Hall (1990)). When the truncation lag is chosen according to the value of the t–statistic on the sum of the autoregressive coefficients (e.g. choosing k such that this t–statistic is minimized), no asymptotic distribution results are yet available. In particular the asymptotic distribution will not be that tabulated in Fuller, and inappropriate inference may follow from such a procedure.

[7]   The fact that λ needs to be restricted to a subset of the interval (0, 1) is because the behavior of the statistic exhibits excessive variability when the endpoints are included. Choosing a "rectangular window", such as $(\epsilon, 1 - \epsilon)$ for any $\epsilon > 0$, is one way to get rid of the problem caused by the endpoints (choosing $\epsilon = 0$ would imply that $t_{\hat{\alpha}}(i,T_B^*,\bar{k})$ is unbounded as T increases, see Andrews (1990)). However, this is not the only way to downweight the influence of the endpoints. An alternative is to take the minimum of the t–statistics over the full interval (0,1) and weighting the statistic by $\lambda(1-\lambda)$. This, however, implies a different asymptotic distribution. See Deshayes and Picard (1986) for a more thorough discussion of these issues.

[8]   In the simulations, kmax was selected to be 5 mainly for computational reasons. The processing time becomes excessive when greater values are used. In some instances, we performed some robustness check by allowing a greater value of kmax. The results were very similar. In practice, kmax can be specified to be greater than 5. What is important is that it be greater than the true order of the autoregression. In principle, the greater kmax is the more likely this criterion will be fulfilled. However, this must be balanced by the fact that estimating high order autoregressions may involve the presence of substantial multicolinearity, making difficult the interpretation of the results.

[9]   We also tabulated critical values for the following additional cases. When k is fixed,

we considered values up to k = 10, the results showed no interesting differences and are not reported. In the case where k is chosen using a test of significance, we also tabulated critical values when the size of the test is 5%. The results, available upon request, are very similar to the 10% size case.

[10] This, however, presumes using appropriate finite sample critical values as opposed to the commonly used asymptotic critical values. No studies are yet available investigating whether this feature is present in the standard Dickey–Fuller (1979) testing procedure.

[11] Given that the unit root is rejected for both series, one could test whether the change in mean is significant using the t–statistic for testing $\delta = 0$ in the first step regression. To carry proper asymptotic inferences would require, however, correcting the standard errors to account for serial correlation. Such a test is easier to perform in the "innovational outlier model" where no such correction is necessary given that the serial correlation is parametrically taken into account in the same regression.

[12] The choice of these countries was determined by data availability. We used data from Dornbusch and Vogelsang (1990), Lee (1976) and the IMF (1988). All series covered the period of 1900–1987.

[13] Using the standard Dickey–Fuller tests, we were not able to reject the unit root hypothesis for any of the series except for Finland/Japan which can be rejected at the 1% significance level. Once we accounted for the shift in mean, we could reject the unit root hypothesis at the 5% significance level in all cases using the "innovational outlier model". This was true regardless of the procedure used to pick the break year T_B or the lag length k.

[14] Once again, the choice of countries was determined by data availability. The sources were Dornbusch and Vogelsang (1990), Lee (1976) and the IMF (1988). The period for France and Sweden was 1900–1988 and that for Canada and Italy was 1914–1987.

[15] Using standard Dickey–Fuller tests we were able to reject the unit root hypothesis at the 1% level for Finland/Canada and Finland/France, at the 5% level for Finland/Sweden and at the 10% level for Finland/Italy.

REFERENCES

Andrews, D.W.K. (1990): "Tests for Parameter Instability and Structural Change with Unknown Change Point," Cowles Foundation Discussion Paper No. 943, Yale University.

Banerjee, A., R.L. Lumsdaine, and J.H. Stock (1990): "Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis," mimeo, Harvard University.

Bank of Finland (1947): "Monthly Bulletin," vol. 21, Nos 4–6.

Board of Governors of the Federal Reserve (1943): *Banking and Monetary Statistics, 1914–1940*, Washington: National Capital Press.

Board of Governors of the Federal Reserve (1976): *Banking and Monetary Statistics, 1941–1970*, Washington: National Capital Press.

Christiano, L.J. (1988): "Searching for Breaks in GNP," National Bureau of Economic Research Working Paper No. 2695.

Corbae, D. and S. Ouliaris (1988): "Cointegration and Tests of PPP," *Review of Economics and Statistics* 4, 508–511.

Davies, R.B. (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternative," *Biometrika* 74, 33–43.

Deshayes, J. and D. Picard (1986): "Off–line Statistical Analysis of Change–point Models Using Non–parametric and Likelihood Methods," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste (eds), New York: Springer–Verlag, 103–168.

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.

Dornbusch, R. (1988): *Exchange Rates and Inflation*, Cambridge: MIT Press.

Dornbusch, R and T.J. Vogelsang (1990): "Real Exchange Rates and Purchasing Power Parity," in *Trade Theory and Economic Reform : North, South and East, Essays in Honor of Bela Balassa*, J. de Melo and A. Sapir (eds).

Enders, W. (1988): "ARIMA and Cointegration–Tests of Purchasing Power Parity," *Review of Economics and Statistics* 4, 504–508.

Enders, W. (1989): "Unit Roots and the Real Exchange Rate Before World War I : The Case of Britain and the USA," *Journal of International Money and Finance* 8, 59–73.

Friedman, M. and A.J. Schwartz (1982): *Monetary Trends in the United States and the United Kingdom : Their Relation to Income, Prices and Interest Rates, 1867–1975*, Chicago: The University of Chicago Press.

Fuller, W.A. (1976): *Introduction to Statistical Time Series*, New York: John Wiley and Sons.

Hall, A. (1990): "Testing for a Unit Root in Time Series With Pretest Data Based Model Selection," mimeo, North Carolina State University.

International Monetary Fund (1988): *International Financial Statistics: Yearbook.*

Jones, R. W. (1979): *International Trade : Essays in Theory*, New York: North–Holland.

Lee, L. H. (1976): *Purshasing Power Parity*, New York: Marcel Dekker.

Mitchell, B.R. (1981): *European Historical Statistics, 1750–1975*, New–York: Columbia University Press.

Nelson, C.R. and C.I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics* 10, 139–162.

Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica* 57, 1361–1401.

Perron, P. (1990a): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics* 8, 153–162.

Perron, P. (1990b): "Test Consistency with Varying Sampling Frequency," forthcoming in *Econometric Theory*.

Perron, P. (1990c): "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," Econometric Research Program Memorandum No. 350, Princeton University.

Press, W.H., B.P. Flanery, S.A. Teukolsky, and W.T. Vetterling (1986): *Numerical Recipes: The Art of Scientific Computing*, Cambridge: Cambridge University Press.

Said, S.E., and D.A. Dickey (1984): "Testing for Unit Roots in Autoregressive–Moving Average of Unknown Order," *Biometrika* 71, 599–608.

Taylor, M. and P. McMahon (1988): "Long–run PPP in the 1920s," *European Economic Review* 32, 179–197.

Tiao, G.C. (1985): "Autoregressive Moving Average Models, Intervention Problems and Outlier Detection in Time Series," in *Handbook of Statistics* 5, E.J. Hannan, P.R. Krishnaiah, and M.M. Rao (eds), Amsterdam: North–Holland, 85–118.

U. S. Bureau of Census (1976): *The Statistical History of the United States, From Colonial Times to the Present*, New York: Basic Books.

Zivot, E. and D.W.K. Andrews (1990): "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," Cowles Foundation Discussion Paper No. 944, Yale University.

MATHEMATICAL APPENDIX

In the proof that follows we only show weak convergence of the finite sample distribution without proving tightness. The proof is made complete using arguments in Zivot and Andrews (1990) who show that the statistics can be expressed as continuous mappings from $D[0,1]$ to R^1 provided we take the infimum with respect to λ over a closed subset of the interval $(0,1)$.

Given that the statistics of interest are constructed applying a particular regression for all values of T_B, it will, in general, be the case that the postulated value of the break in that regression does not correspond to the true value of the break in the data–generating process (8). Hence, we cannot appeal to a finite sample invariance with respect to the true value of δ. However, as the proofs below show, such an invariance property holds asymptotically. To make the argument precise, let the true value of the break be denoted by $T_B{}'$. Under the null hypothesis, the data–generating process is :

$$y_t = \delta D(TB')_t + y_{t-1} + e_t = \delta D(TB)_t + y_{t-1} + e_t^* , \tag{A.1}$$

where $D(TB')_t = 1$ if $t = T_B{}' + 1$ and 0 otherwise, $D(TB)_t = 1$ if $t = T_B + 1$ and 0 otherwise, and $e_t^* = e_t + \delta(D(TB')_t - D(TB)_t)$. Alternatively, we can express y_t as :

$$y_t = y(0) + \delta DU'_t + S_t ; \tag{A.2}$$

where $S_t = \Sigma_{j=1}^t e_t$, $DU'_t = 0$ if $t \leq T_B{}'$ and $DU'_t = 1$ if $t > T_B{}'$.

Proof of (9) : We prove (9) for the "additive outlier model" only in the case where no additional lags are included in regression (4) and the data–generating process is given by (8). As mentioned in the text, we rely on arguments made in Said and Dickey (1984) when additional correlation is present and higher order autoregressions are estimated. The result for the "innovational outlier model" follows, with appropriate modifications using results in Zivot and Andrews (1990), and details for this case are omitted. Consider first applying regression (3) for a given value T_B. We assume $T_B \leq T_B{}'$ without loss of generality given that the asymptotic distribution is invariant to $T_B{}'$. Throughout, we denote $\lambda = T_B/T$

and $\lambda' = T_B'/T$. Simple algebra yields :

$$\tilde{y}_t = y_t - \bar{Y}_a = S_t - \bar{S}_a \qquad \text{if } t \leq T_B \text{ ,}$$

$$\tilde{y}_t = y_t - \bar{Y}_b = S_t - \bar{S}_b - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B < t \leq T_B' \text{ ,} \qquad \text{(A.3)}$$

$$\tilde{y}_t = y_{t*} - \bar{Y}_b = S_t - \bar{S}_b + \delta - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B' < t \leq T \text{ ;}$$

with $\bar{Y}_a = T_B^{-1}\Sigma_1^{T_B}y_t = \lambda^{-1}T^{-1}\Sigma_1^{T_B}y_t$, $\bar{Y}_b = (T-T_B)^{-1}\Sigma_{T_B+1}^{T}y_t = (1-\lambda)^{-1}$
$T^{-1}\Sigma_{T_B+1}^{T}y_t$, $\bar{S}_a = T_B^{-1}\Sigma_1^{T_B}S_t = \lambda^{-1}T^{-1}\Sigma_1^{T_B}S_t$, and $\bar{S}_b = (T-T_B)^{-1}\Sigma_{T_B+1}^{T}S_t =$
$(1-\lambda)^{-1}T^{-1}\Sigma_{T_B+1}^{T}S_t$. Denote the residuals from a projection of $\tilde{y}_t$ on $D(TB)_t$ by $\tilde{y}_t^*$, and
similarly the residuals from a projection of $\tilde{y}_{t-1}$ on $D(TB)_t$ by $\tilde{y}_{t-1}^{**}$. We have :

$$\tilde{y}_t^* = S_t - \bar{S}_a \qquad \text{if } t \leq T_B \text{ ,}$$

$$\tilde{y}_t^* = 0 \qquad \text{if } t = T_B + 1 \text{ ,}$$

$$\tilde{y}_t^* = S_t - \bar{S}_b - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B + 1 < t \leq T_B' \text{ ,} \qquad \text{(A.4)}$$

$$\tilde{y}_t^* = S_t - \bar{S}_b + \delta - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B' < t \leq T \text{ .}$$

Similarly,

$$\tilde{y}_{t-1}^{**} = S_{t-1} - \bar{S}_a \qquad \text{if } t \leq T_B \text{ ,}$$

$$\tilde{y}_{t-1}^{**} = 0 \qquad \text{if } t = T_B + 1 \text{ ,}$$

$$\tilde{y}_{t-1}^{**} = S_{t-1} - \bar{S}_b - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B + 1 < t \leq T_B' \text{ ,} \qquad \text{(A.5)}$$

$$\tilde{y}_{t-1}^{**} = S_{t-1} - \bar{S}_b + \delta - \delta(1-\lambda')/(1-\lambda) \qquad \text{if } T_B' < t \leq T \text{ .}$$

Using (A.4) and (A.5), we deduce that :

$$\tilde{y}_t^* - \tilde{y}_{t-1}^{**} = e_t \qquad \text{if } t \neq T_B + 1 \text{ and } t \neq T_B' + 1 \text{ ,}$$

$$= 0 \qquad \text{if } t = T_B + 1 , \qquad\qquad (A.6)$$

$$= e_{T_B'+1} + \delta \quad \text{if } t = T_B' + 1.$$

The t–statistic for testing $\alpha = 1$ in regression (4) with $k = 0$ is given by :

$$t_{\hat{\alpha}} = (\hat{\alpha} - 1)[\Sigma_2^T (\tilde{y}_{t-1}^*)^2]^{1/2}/\hat{s} = T^{-1}\Sigma_2^T \tilde{y}_{t-1}^{**}(\tilde{y}_t^* - \tilde{y}_{t-1}^{**})/[\hat{s}^2 T^{-2}\Sigma_2^T (\tilde{y}_{t-1}^{**})^2]^{1/2} , \qquad (A.7)$$

where $\hat{s}^2 = T^{-1}\Sigma_2^T \hat{u}_t^2$ with $\hat{u}_t = \tilde{y}_t - \hat{\omega}D(TB)_t - \hat{\alpha}\tilde{y}_{t-1}$. Consider first the numerator in (A.7). Simple manipulations yields :

$$T^{-1}\Sigma_2^T \tilde{y}_{t-1}^{**}(\tilde{y}_t^* - \tilde{y}_{t-1}^{**}) = T^{-1}\Sigma_2^T S_{t-1}e_t - (T^{-1/2}\Sigma_1^{T_B} e_t)\lambda^{-1}T^{-3/2}\Sigma_1^{T_B} S_t$$

$$- (T^{-1/2}\Sigma_{T_B+1}^T e_t)(1 - \lambda)^{-1}T^{-3/2}\Sigma_{T_B+1}^T S_t + o_p(1) .$$

Using Lemma A.3 of Perron (1989), we deduce that :

$$T^{-1}\Sigma_2^T \tilde{y}_{t-1}^{**}(\tilde{y}_t^* - \tilde{y}_{t-1}^{**}) \Rightarrow \sigma^2\Big\{(1/2)(W(1)^2 - 1) - \lambda^{-1}W(\lambda)\int_0^\lambda W(r)dr$$

$$- (1 - \lambda)^{-1}[W(1) - W(\lambda)][\int_\lambda^1 W(r)dr]\Big\}$$

$$\equiv \sigma^2 \int_0^1 W^*(\lambda,r)dW(r) . \qquad (A.8)$$

Consider now the denominator in (A.7). We have, using (A.4) and some manipulations :

$$T^{-2}\Sigma_2^T (\tilde{y}_{t-1}^{**})^2 = T^{-2}\Sigma_2^{T_B}(S_{t-1} - \bar{S}_a)^2 + T^{-2}\Sigma_{T_B+1}^{T_B'}(S_{t-1} - \bar{S}_b - \delta(1 - \lambda')/(1 - \lambda))^2$$

$$+ T^{-2}\Sigma_{T_B'+1}^T (S_{t-1} - \bar{S}_b + \delta - \delta(1 - \lambda')/(1 - \lambda))^2$$

$$= T^{-2}\Sigma_2^T S_{t-1}^2 - \lambda^{-1}T^{-3}(\Sigma_1^{T_B} S_t)^2 - (1 - \lambda)^{-1}T^{-3}(\Sigma_{T_B+1}^T S_{t-1})^2 + o_p(1)$$

$$\Rightarrow \sigma^2\Big\{\int_0^1 W(r)^2 dr - \lambda^{-1}(\int_0^\lambda W(r)dr)^2 - (1 - \lambda)^{-1}(\int_\lambda^1 W(r)dr)^2\Big\}$$

$$\equiv \sigma^2 \int_0^1 W^*(\lambda,r)^2 dr .\qquad\qquad (A.9)$$

Tedious but straightforward manipulations show that $\hat{s}^2 \to \sigma^2$ in probability. The result in (9) is then proved using (A.7) through (A.9).

Remark : Perron (1990a) considered testing for a unit root in the "additive outlier model" without the regressors $D(TB)_t$ in regression (4) and claimed that the asymptotic distribution of the t−statistic for testing $\alpha = 1$ was the same as the expression in (9). The above results show this claim to be erroneous. To see this note that, in this case, the t−statistic is :

$$t_{\hat{\alpha}} = T^{-1}\Sigma_2^T \tilde{y}_{t-1}(\tilde{y}_t - \tilde{y}_{t-1})/[\hat{s}^2 T^{-2}\Sigma_2^T(\tilde{y}_{t-1})^2]^{1/2}, \qquad\qquad (A.10)$$

where $\hat{s}^2 = T^{-1}\Sigma_2^T(\tilde{y}_t - \hat{\alpha}\tilde{y}_{t-1})^2$. Using (A.3), we deduce after some algebra that :

$$T^{-1}\Sigma_2^T \tilde{y}_{t-1}(\tilde{y}_t - \tilde{y}_{t-1}) = T^{-1}\Sigma_2^T S_{t-1}e_t - T^{-1}S_T\bar{S}_b + T^{-1}(\bar{S}_b - \bar{S}_a)\bar{S}_a + o_p(1)$$

$$\Rightarrow \sigma^2\Big\{(1/2)(W(1)^2 - 1) - (1-\lambda)^{-1}W(1)\int_\lambda^1 W(r)dr$$

$$+ \lambda^{-1}\int_0^\lambda W(r)dr[(1-\lambda)^{-1}\int_\lambda^1 W(r)dr - \lambda^{-1}\int_0^\lambda W(r)dr]\Big\} \quad (A.11)$$

$$T^{-2}\Sigma_2^T \tilde{y}_{t-1}^2 = T^{-2}\Sigma_2^T S_{t-1}^2 - \lambda^{-1}T^{-3}(\Sigma_1^{T_B}S_t)^2 - (1-\lambda)^{-1}T^{-3}(\Sigma_{T_B+1}^T S_{t-1})^2 + o_p(1)$$

$$\Rightarrow \sigma^2\Big\{\int_0^1 W(r)^2 dr - \lambda^{-1}(\int_0^\lambda W(r)dr)^2 - (1-\lambda)^{-1}(\int_\lambda^1 W(r)dr)^2\Big\}$$

$$\equiv \sigma^2 \int_0^1 W^*(\lambda,r)^2 dr ; \qquad\qquad (A.12)$$

$$\hat{s}^2 = T^{-1}\Sigma_2^T e_t^2 + T^{-1}(\bar{S}_a - \bar{S}_b)^2 + o_p(1)$$

$$\Rightarrow \sigma^2[1 + (\lambda^{-1}\int_0^\lambda W(r)dr - (1-\lambda)^{-1}\int_\lambda^1 W(r)dr)^2] . \qquad\qquad (A.13)$$

(A.10) through (A.13) show the asymptotic distribution of the t−statistic $t_{\hat{\alpha}}$ when $D(TB)_t$ is not introduced as a regressor in (4) to be different from that stated in (9) and in Perron (1990a).

Proof of (12) : To prove (12), we need only show that the limiting distribution of $t_{\hat{\delta}}$ is $Z(\lambda)$ when $t_{\hat{\delta}}$ is the t−statistic for $\delta = 0$ in regression (7) with $k = 0$ under the null hypothesis that y_t is generated by (8). As in Perron (1989, 1990a), the presence of the regressor $D(TB)_t$ in (7) does not change the asymptotic distribution of the t−statistic. To simplify notation, we therefore omit this regressor in the following derivations and express the t−statistic as :

$$t_{\hat{\delta}} = \left\{ [Z'M_X Z]^{-1}[Z'M_X E] \right\}_{11} \left\{ \hat{s}^2 [(Z'M_X Z)^{-1}]_{11} \right\}^{-1/2} + o_p(1) ; \qquad (A.14)$$

where $Z = [DU_t \; y_{t-1}]$ [(T−1) x 2], $X' = [1, 1, ..., 1]$ [1 x (T−1)], $E = [e_2^*, e_3^*, ..., e_T^*]$ [1 x (T−1)], $M_X = I - X(X'X)^{-1}X'$ and $\hat{s}^2 = T^{-1}\Sigma_2^T \hat{u}_t^2$ with $\hat{u}_t$ the regression residuals from applying OLS to (7) and e_t^* defined in (A.1). Note that $\hat{s}^2 \to \sigma^2$ (in probability), see Perron (1990a). Straightforward algebra yields :

$$[(Z'M_X Z)^{-1}]_{11} = [\Sigma_2^T y_{t-1}^2 - T^{-1}(\Sigma_2^T y_{t-1})^2]/J \qquad (A.15)$$

where $J = \lambda(1 - \lambda)T[\Sigma_2^T y_{t-1}^2 - T^{-1}(\Sigma_2^T y_{t-1})^2] - [\lambda\Sigma_2^T y_{t-1} - \Sigma_2^{T_B} y_{t-1}]^2$.

Note that, using Lemma A.3 of Perron (1989) :

$$T^{-3}J \Rightarrow \sigma^2\lambda(1 - \lambda)[\int_0^1 W(r)^2 dr - (\int_0^1 W(r)dr)^2] - \sigma^2[\lambda\int_0^1 W(r)dr - \int_0^\lambda W(r)dr]^2$$

$$= \sigma^2\lambda(1 - \lambda)A - \sigma^2 B(\lambda)^2 \equiv \sigma^2 K(\lambda)/A . \qquad (A.16)$$

Also :

$$T[(Z'M_X Z)^{-1}]_{11} \Rightarrow \sigma^2 A/[K(\lambda)/A] = \sigma^2 A^2/K(\lambda) ;$$

using (A.15) and (A.16). Finally, simple algebra, using in particular (A.1), shows that

$$T^{1/2}\left\{ [Z'M_X Z]^{-1}[Z'M_X E] \right\}_{11} =$$

$$(1/T^{-3}J)\Big\{[T^{-2}\Sigma_2^T y_{t-1}^2 - T^{-3}(\Sigma_2^T y_{t-1})^2]\,[T^{-1/2}\Sigma_{T_B+1}^T e_t^* - (1-\lambda)T^{-1/2}\Sigma_2^T e_t^*]$$

$$-[\lambda T^{-3/2}\Sigma_2^T y_{t-1} - T^{-3/2}\Sigma_2^{T_B} y_{t-1}]\,[T^{-1}\Sigma_2^T y_{t-1}e_t^* - T^{-3/2}\Sigma_2^T y_{t-1}T^{-1/2}\Sigma_2^T e_t^*]\Big\}\ .$$

It is straightforward to see that the asymptotic distributions of $T^{-1/2}\Sigma_{T_B+1}^T e_t^*$, $T^{-1/2}\Sigma_2^T e_t^*$ and $T^{-1}\Sigma_2^T y_{t-1}e_t^*$ are identical to those of $T^{-1/2}\Sigma_{T_B+1}^T e_t$, $T^{-1/2}\Sigma_2^T e_t$ and $T^{-1}\Sigma_2^T y_{t-1}e_t$. Using this fact, (A.16) and Lemma A.3 of Perron (1989), we have :

$$T^{1/2}\Big\{[Z'M_X Z]^{-1}[Z'M_X E]\Big\}_{11} \Rightarrow$$

$$\sigma^2(A/K(\lambda))\Big\{A[\lambda W(1) - W(\lambda)] - B(\lambda)[(1/2)(W(1)^2 - 1) - W(1)\int_0^1 W(r)dr]\Big\}$$

$$\equiv \sigma^2 AH(\lambda)/K(\lambda)\ . \tag{A.17}$$

Using (A.14) through (A.17) we deduce, as required, that :

$$t_{\hat{\delta}} \Rightarrow [\sigma^2 AH(\lambda)/K(\lambda)]/[\sigma^4 A^2/K(\lambda)]^{1/2} = H(\lambda)/K(\lambda)^{1/2}\ .$$

Proof of (13) : To prove (13) we need only show that the limiting distribution of $T^{-1/2}t_{\hat{\delta}}$ is $Q(\lambda)$ when $t_{\hat{\delta}}$ is the t–statistic for $\delta = 0$ in regression (3) and y_t is generated by (8). The t–statistic is not invariant, in finite samples, with respect to the value of δ. However, the following proof shows that its asymptotic distribution is invariant. First note that :

$$t_{\hat{\delta}} = \hat{\delta}/[\hat{s}^2(X'X)_{22}^{-1}]^{1/2}\ , \tag{A.18}$$

where $(X'X)_{22}^{-1} = T/[T_B(T - T_B)] = [T\lambda(1-\lambda)]^{-1}$, $\hat{s}^2 = T^{-1}\Sigma_{t=1}^T \hat{u}_t^2 = T^{-1}\Sigma_{t=1}^T(y_t - \hat{\mu} - \hat{\delta}DU_t)^2$ where $\hat{\mu} = (\lambda T)^{-1}\Sigma_{t=1}^{T_B} y_t$ and

$$\hat{\delta} = [(1-\lambda)T]^{-1}\Sigma_{t=T_B+1}^T y_t - [\lambda T]^{-1}\Sigma_{t=1}^{T_B} y_t\ . \tag{A.19}$$

Using (A.2), we can write (after some rearrangements) :

$$\hat{\delta} - \delta = [(1-\lambda)T]^{-1}\Sigma_{T_B+1}^{T}S_t - [\lambda T]^{-1}\Sigma_1^{T_B}S_t - \delta(1-\lambda')/(1-\lambda) . \tag{A.20}$$

Using Lemma A.3 of Perron (1989) we deduce, from (A.20), that :

$$T^{-1/2}(\hat{\delta} - \delta) \Rightarrow \sigma\Big\{(1-\lambda)^{-1}\int_{\lambda}^{1}W(r)dr - \lambda^{-1}\int_0^{\lambda}W(r)dr\Big\}$$

$$= \sigma[(1-\lambda)\lambda]^{-1}\Big\{\lambda\int_0^1W(r)dr - \int_0^{\lambda}W(r)dr\Big\}$$

$$\equiv \sigma[(1-\lambda)\lambda]^{-1}B(\lambda). \tag{A.21}$$

Note that (A.21) implies that $T^{-1/2}(\hat{\delta}-\delta)$ has the same asymptotic distribution as $T^{-1/2}\hat{\delta}$ since $T^{-1/2}\delta \to 0$ as $T \to \infty$. Using similar arguments,

$$\hat{s}^2 = T^{-1}\Sigma_{t=1}^{T}y_t^2 - [\lambda T^2]^{-1}(\Sigma_{t=1}^{T_B}y_t)^2 - [(1-\lambda)T^2]^{-1}(\Sigma_{t=T_B+1}^{T}y_t)^2$$

$$= T^{-1}\Sigma_{t=1}^{T}S_t^2 - [\lambda T^2]^{-1}(\Sigma_{t=1}^{T_B}S_t)^2 - [(1-\lambda)T^2]^{-1}(\Sigma_{t=T_B+1}^{T}S_t)^2 + o_p(T) ;$$

and

$$T^{-1}\hat{s}^2 \Rightarrow \sigma^2\Big\{\int_0^1W(r)^2dr - \lambda^{-1}(\int_0^{\lambda}W(r)dr)^2 - (1-\lambda)^{-1}(\int_{\lambda}^1W(r)dr)^2\Big\}$$

$$= \sigma^2[(1-\lambda)\lambda]^{-1}\Big\{(1-\lambda)\lambda\int_0^1W(r)^2dr - (1-\lambda)(\int_0^{\lambda}W(r)dr)^2 - \lambda(\int_{\lambda}^1W(r)dr)^2\Big\}$$

$$= \sigma^2[(1-\lambda)\lambda]^{-1}K(\lambda)/A , \text{ after rearrangements.} \tag{A.22}$$

Using (A.18), (A.21) and (A.22) we obtain :

$$T^{-1/2}t_{\hat{\delta}} = T^{-1/2}\hat{\delta}/[T^{-1}\hat{s}^2[\lambda(1-\lambda)]^{-1}]^{1/2}$$

$$\Rightarrow \sigma[(1-\lambda)\lambda]^{-1}B(\lambda)/[\sigma^2[(1-\lambda)\lambda]^{-2}K(\lambda)/A]^{1/2}$$

$$= B(\lambda)/[K(\lambda)/A]^{1/2} ; \text{ as required.}$$

Table 1 : Percentage Points of the Distribution of $t_{\hat{\alpha}}(AO,T_B^*,k)$; Additive Outlier Model.

Choosing T_B minimizing $t_{\hat{\alpha}}$ in (4).

		1.0 %	2.5 %	5.0 %	10.0 %	90.0 %	95.0 %	97.5 %	99.0 %
T = 50	k = 0	-5.12	-4.77	-4.46	-4.12	-2.27	-2.07	-1.89	-1.69
	k = 2	-5.04	-4.67	-4.35	-4.01	-2.12	-1.96	-1.78	-1.56
	k = 5	-5.04	-4.63	-4.23	-3.91	-2.09	-1.87	-1.71	-1.46
	k = k*	-5.57	-5.20	-4.96	-4.61	-2.69	-2.46	-2.26	-2.04
	k = k(F)	-5.37	-5.04	-4.70	-4.39	-2.46	-2.21	-2.02	-1.86
	k = k(t)	-5.44	-5.12	-4.86	-4.49	-2.55	-2.30	-2.12	-1.97
T = 100	k = 0	-5.07	-4.75	-4.43	-4.12	-2.36	-2.15	-1.96	-1.73
	k = 2	-4.97	-4.56	-4.30	-4.00	-2.26	-2.04	-1.82	-1.62
	k = 5	-4.87	-4.48	-4.19	-3.89	-2.20	-1.97	-1.74	-1.53
	k = k*	-5.46	-5.04	-4.71	-4.42	-2.64	-2.39	-2.19	-1.96
	k = k(F)	-5.24	-4.87	-4.59	-4.28	-2.43	-2.19	-1.99	-1.65
	k = k(t)	-5.30	-4.99	-4.67	-4.35	-2.52	-2.28	-2.12	-1.90
T = 150	k = 0	-5.08	-4.76	-4.43	-4.14	-2.39	-2.17	-1.95	-1.65
	k = 2	-4.88	-4.60	-4.35	-4.08	-2.31	-2.03	-1.87	-1.61
	k = 5	-5.05	-4.58	-4.29	-3.96	-2.23	-1.98	-1.80	-1.59
	k = k*	-5.44	-4.99	-4.73	-4.42	-2.58	-2.32	-2.11	-1.87
	k = k(F)	-5.37	-4.87	-4.56	-4.24	-2.42	-2.15	-1.94	-1.72
	k = k(t)	-5.40	-4.94	-4.68	-4.32	-2.48	-2.20	-1.98	-1.79
T = ∞		-4.95	-4.69	-4.44	-4.19	-2.51	-2.28	-2.10	-1.85

Table 2 : Percentage Points of the Distribution of $t_{\hat{\alpha}}(IO, T_B^*, k)$; Innovational Outlier Model Choosing T_B minimizing $t_{\hat{\alpha}}$ in (7).

		1.0 %	2.5 %	5.0 %	10.0 %	90.0 %	95.0 %	97.5 %	99.0 %
T = 50	k = 0	-5.07	-4.70	-4.37	-4.04	-2.19	-1.99	-1.82	-1.59
	k = 2	-5.10	-4.72	-4.42	-4.04	-2.12	-1.89	-1.68	-1.42
	k = 5	-5.19	-4.79	-4.42	-4.07	-2.06	-1.77	-1.48	-1.10
	k = k*	-5.58	-5.23	-4.93	-4.56	-2.56	-2.31	-2.09	-1.89
	k = k(F)	-5.52	-5.06	-4.81	-4.47	-2.33	-2.04	-1.79	-1.43
	k = k(t)	-5.51	-5.07	-4.76	-4.42	-2.34	-2.09	-1.85	-1.61
T = 100	k = 0	-5.03	-4.71	-4.39	-4.08	-2.33	-2.12	-1.91	-1.68
	k = 2	-5.06	-4.62	-4.39	-4.06	-2.32	-2.04	-1.78	-1.57
	k = 5	-4.94	-4.67	-4.35	-4.11	-2.30	-1.99	-1.78	-1.38
	k = k*	-5.45	-5.00	-4.72	-4.41	-2.59	-2.33	-2.13	-1.86
	k = k(F)	-5.27	-4.88	-4.60	-4.28	-2.43	-2.14	-1.90	-1.54
	k = k(t)	-5.33	-4.88	-4.58	-4.27	-2.41	-2.17	-1.96	-1.68
T = 150	k = 0	-5.07	-4.73	-4.41	-4.11	-2.36	-2.15	-1.93	-1.61
	k = 2	-4.95	-4.68	-4.41	-4.13	-2.34	-2.10	-1.86	-1.65
	k = 5	-5.11	-4.70	-4.43	-4.11	-2.30	-2.04	-1.84	-1.59
	k = k*	-5.36	-4.97	-4.71	-4.40	-2.56	-2.29	-2.07	-1.84
	k = k(F)	-5.21	-4.87	-4.58	-4.24	-2.41	-2.13	-1.89	-1.63
	k = k(t)	-5.17	-4.84	-4.58	-4.24	-2.40	-2.16	-1.88	-1.64
T = ∞		-4.95	-4.69	-4.44	-4.19	-2.51	-2.28	-2.10	-1.85

Table 3 : Percentage Points of the Distribution of $t_{\hat\alpha}(AO,T_B(\hat\delta),k)$; Additive Outlier Model.

Choosing T_B minimizing $t_{\hat\delta}$ in (3).

		1.0 %	2.5 %	5.0 %	10.0 %	90.0 %	95.0 %	97.5 %	99.0 %
T = 50	k = 0	-4.25	-3.86	-3.54	-3.19	-0.84	-0.43	-0.15	0.22
	k = 2	-4.08	-3.76	-3.41	-3.02	-0.69	-0.36	-0.05	0.26
	k = 5	-3.90	-3.46	-3.19	-2.83	-0.72	-0.34	-0.09	0.17
	k = k*	-4.59	-4.25	-3.91	-3.56	-1.17	-0.84	-0.50	-0.19
	k = k(F)	-4.33	-3.94	-3.62	-3.28	-0.76	-0.35	-0.09	0.22
	k = k(t)	-4.48	-4.08	-3.72	-3.36	-0.87	-0.49	-0.17	0.21
T = 100	k = 0	-4.11	-3.85	-3.55	-3.19	-0.80	-0.46	-0.15	0.27
	k = 2	-4.05	-3.78	-3.45	-3.16	-0.75	-0.38	-0.12	0.21
	k = 5	-4.10	-3.65	-3.39	-3.05	-0.73	-0.34	-0.07	0.20
	k = k*	-4.57	-4.14	-3.89	-3.55	-1.00	-0.71	-0.41	-0.12
	k = k(F)	-4.42	-3.98	-3.66	-3.30	-0.78	-0.43	-0.06	0.23
	k = k(t)	-4.42	-4.03	-3.70	-3.32	-0.80	-0.50	-0.14	0.22
T = 150	k = 0	-4.10	-3.80	-3.51	-3.16	-0.81	-0.40	-0.01	0.28
	k = 2	-4.08	-3.72	-3.43	-3.10	-0.75	-0.40	0.06	0.37
	k = 5	-4.09	-3.66	-3.39	-3.06	-0.73	-0.32	0.01	0.37
	k = k*	-4.42	-4.11	-3.79	-3.42	-0.98	-0.61	-0.20	0.12
	k = k(F)	-4.28	-3.91	-3.58	-3.24	-0.76	-0.35	-0.02	0.37
	k = k(t)	-4.32	-3.95	-3.65	-3.26	-0.79	-0.38	-0.04	0.28
T = ∞		-4.18	-3.90	-3.61	-3.27	-0.79	-0.39	-0.07	0.25

Table 4 : Percentage Points of the Distribution of $t_{\hat{\alpha}}(IO,T_B(\hat{\delta}),k)$; Innovational Outlier Model. Choosing T_B minimizing $t_{\hat{\delta}}$ in (7).

		1.0 %	2.5 %	5.0 %	10.0 %	90.0 %	95.0 %	97.5 %	99.0 %
T = 50	k = 0	-4.68	-4.32	-3.97	-3.63	-0.36	0.12	0.53	0.86
	k = 2	-4.93	-4.36	-4.01	-3.62	-0.16	0.41	0.85	1.32
	k = 5	-4.84	-4.31	-3.96	-3.63	0.17	0.81	1.35	1.80
	k = k*	-5.27	-4.78	-4.42	-3.99	0.26	0.89	1.37	1.80
	k = k(F)	-5.14	-4.74	-4.31	-3.86	0.15	0.84	1.34	1.72
	k = k(t)	-5.10	-4.70	-4.26	-3.82	-0.02	0.69	1.18	1.60
T = 100	k = 0	-4.74	-4.39	-4.05	-3.73	-0.42	0.09	0.44	1.05
	k = 2	-4.65	-4.31	-4.06	-3.72	-0.23	0.24	0.68	1.24
	k = 5	-4.75	-4.37	-4.03	-3.69	-0.03	0.52	1.03	1.42
	k = k*	-5.03	-4.66	-4.32	-3.95	0.09	0.60	1.09	1.58
	k = k(F)	-4.95	-4.58	-4.22	-3.84	-0.10	0.43	0.89	1.33
	k = k(t)	-4.91	-4.49	-4.19	-3.81	-0.21	0.41	0.89	1.33
T = 150	k = 0	-4.84	-4.41	-4.12	-3.77	-0.29	0.14	0.50	0.85
	k = 2	-4.75	-4.45	-4.17	-3.75	-0.27	0.20	0.60	0.96
	k = 5	-4.85	-4.45	-4.16	-3.75	-0.29	0.36	0.77	1.16
	k = k*	-5.05	-4.70	-4.38	-3.96	-0.13	0.49	0.88	1.22
	k = k(F)	-4.97	-4.62	-4.27	-3.86	-0.23	0.35	0.76	1.11
	k = k(t)	-4.97	-4.62	-4.26	-3.86	-0.24	0.37	0.73	1.01
T = ∞		-4.73	-4.44	-4.19	-3.86	-0.42	0.09	0.50	0.93

TABLE 5 : Empirical Results , $t_{\hat{\alpha}}(AO,T_B^*,k)$; Real Exchange Rates ; Additive Outlier Model.

Regression : $y_t = \mu + \delta DU_t + \tilde{y}_t$; $\tilde{y}_t = \omega D(TB)_t + \alpha \tilde{y}_{t-1} + \Sigma_{i=1}^{k} c_i \Delta \tilde{y}_{t-i} + e_t$.

Series	T_B	k	$\hat{\mu}$	$\hat{\delta}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	p-value (k fixed)	p-value (k*)	p-value (k(t))	p-value (k(F))
US/Fin (GDP) (kmax = 5)	1938	1	−0.97 (−33.13)	−.4066 (−10.32)	.581	−5.18	<.01	.02	.01	.01
US/UK (CPI) (kmax = 5)	1943	1	−1.66 (−100.6)	.3529 (14.62)	.686	−4.98	.01	.03	.02	.02

TABLE 6 : Empirical Results, $t_{\hat{\alpha}}(IO, T_B^*, k)$; Real Exchange Rates ; Innovational Outlier Model.

Regression : $y_t = \mu + \delta DU_t + \theta D(TB)_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + e_t$.

Series	T_B	k	$\hat{\mu}$	$\hat{\delta}$	$\hat{\theta}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	p-value (k fixed)	p-value (k*)	p-value (k(t))	p-value (k(F))
US/Fin (GDP) (kmax = 5)	1938	1	-.388 (-4.84)	-.176 (-4.22)	.222 (1.74)	.598	-5.02	<.01	.03		
	1945	1	-.363 (-4.35)	-.150 (-3.60)	0.117 (0.91)	.641	-4.43	.05		.07	.07
US/UK (CPI) (kmax = 5)	1944	1	-.512 (-4.80)	0.112 (4.14)	-.072 (-1.03)	.690	-4.83	.02	.04	.03	.03

TABLE 7 : Empirical Results, $t_{\hat{\alpha}}(AO, T_B(\hat{\delta}), k)$; Real Exchange Rates ; Additive Outlier Model.

Regression : $y_t = \mu + \delta DU_t + \tilde{y}_t$; $\tilde{y}_t = \omega D(TB)_t + \alpha \tilde{y}_{t-1} + \sum_{i=1}^{k} c_i \Delta \tilde{y}_{t-i} + e_t$.

	T_B	$\hat{\delta}$	Min $t_{\hat{\alpha}}$ unrestricted				With t-test on last lag				With F-test on additional lags			
			k	$t_{\hat{\alpha}}$	p-value (fixed k)	p-value (k*)	k	$t_{\hat{\alpha}}$	p-value (fixed k)	p-value (k(t))	k	$t_{\hat{\alpha}}$	p-value (fixed k)	p-value (k(F))
US/Fin (GDP) (kmax = 5)	1940	−.4204 (−11.27)	1	−4.83 (.586)	<.01	<.01	1	−4.83 (.586)	<.01	<.01	1	−4.83 (.586)	<.01	<.01
US/UK (CPI) (kmax = 5)	1946	0.3633 (15.80)	1	−4.34 (.706)	<.01	.02	1	−4.34 (.706)	<.01	.01	1	−4.34 (.706)	<.01	.01

Note : t-statistics are in parentheses except for the columns labelled $t_{\hat{\alpha}}$ where the entries are the estimates of α.

Table 8 : Empirical Results, $t_{\hat{\alpha}}(IO, T_B(\hat{\delta}), k)$; Real Exchange Rates ; Innovational Outlier Model.

Regression : $y_t = \mu + \delta DU_t + \theta D(TB)_t + \alpha y_{t-1} + \Sigma_{i=1}^{k} c_i \Delta y_{t-i} + e_t$.

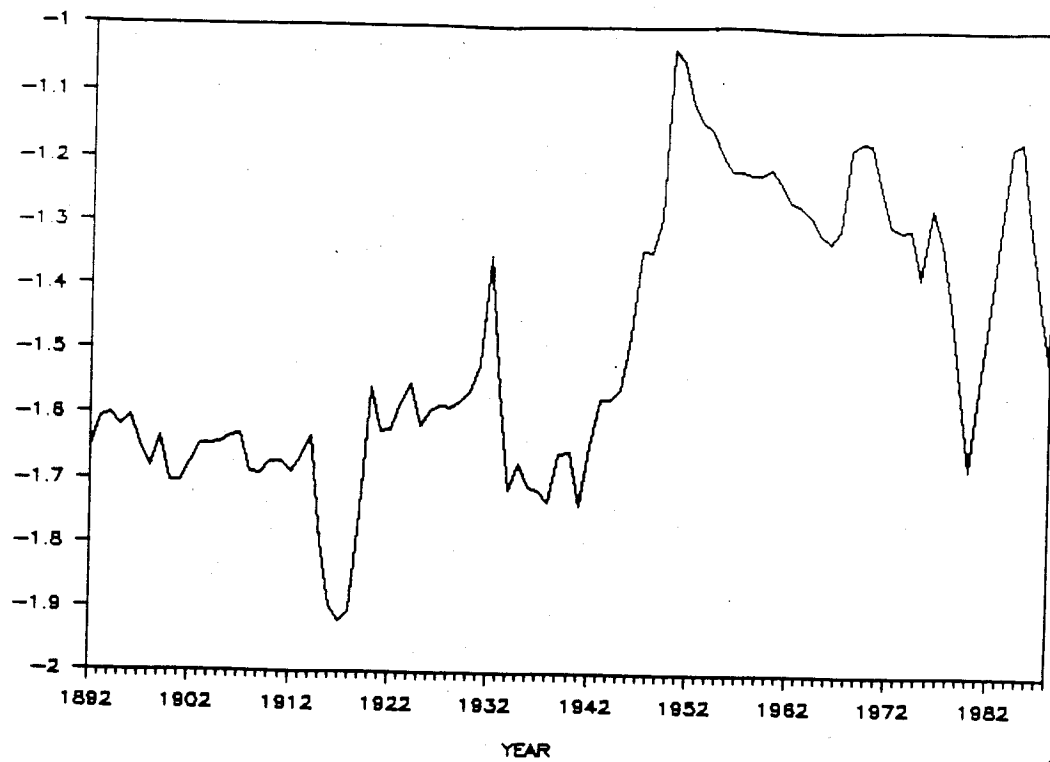| Series	T_B	k	$t_{\hat{\delta}}$	$\hat{\alpha}$	$t_{\hat{\alpha}}$	p-value (k fixed)	p-value (k^*)	p-value ($k(t)$)	p-value ($k(F)$)
US/Fin (GDP) (kmax = 5)	1939	1	-4.24	.589	-4.96	<.01	.02		
	1945	1	-3.60	.641	-4.43	.03		.03	.04
US/UK (CPI) (kmax = 5)	1944	1	4.14	.690	-4.83	<.01	.02	.02	.02

Figure 1(a):

Logarithm of the U.S./U.K. real exchange rate (1892-1988, annual) based on the
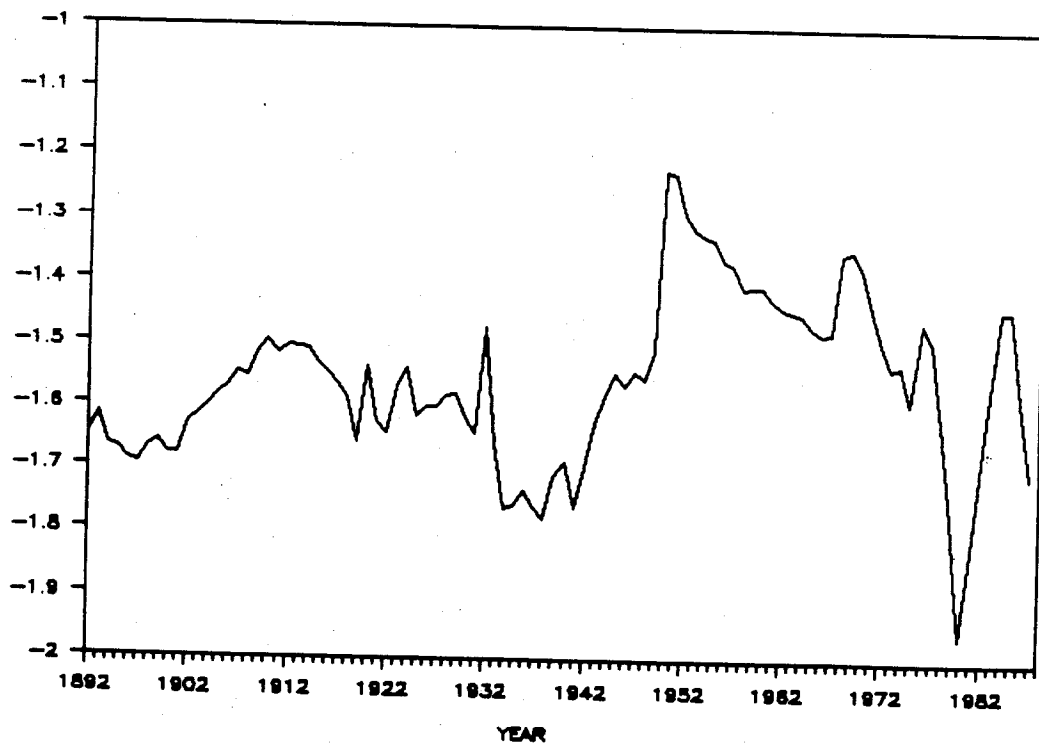consumer price indices (CPI).

Figure 1 (b):

Logarithm of the U.S./U.K. real exchange rate (1892-1988, annual) based on the GNP
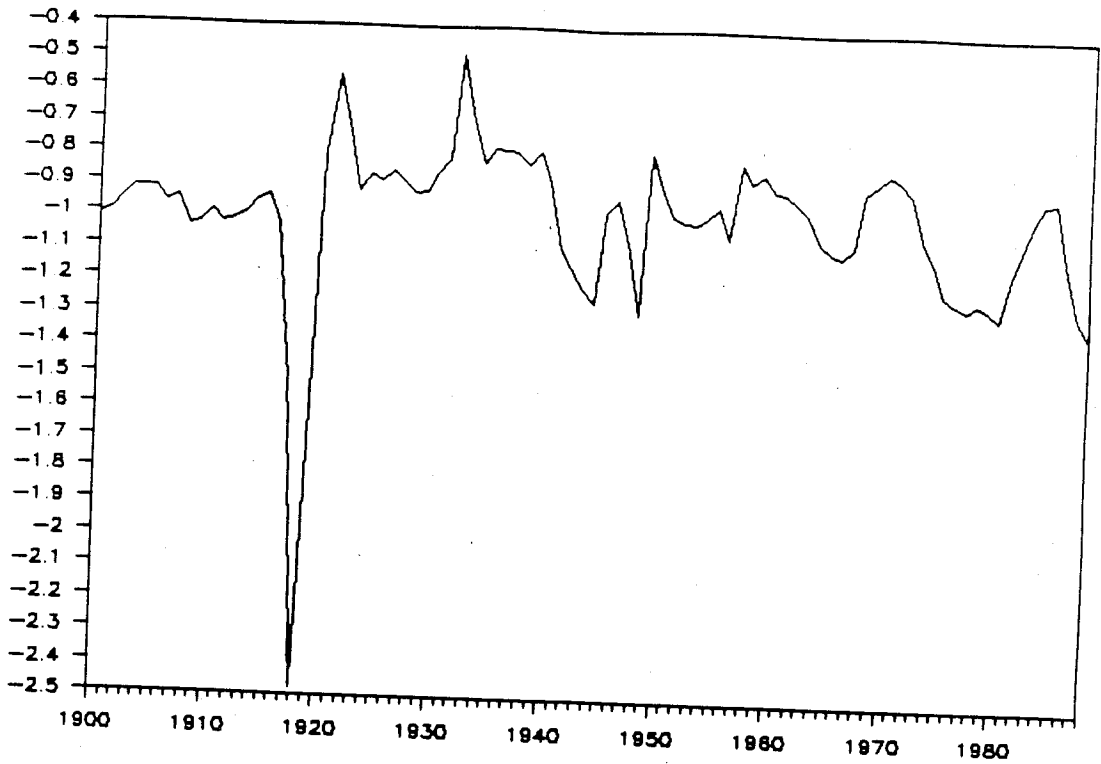deflators as the price indices.

Figure 2 (a):

Logarithm of the U.S./Finland real exchange rate (1900-1987, annual) based on the consumer price indices (CPI).

Figure 2 (b):

Logarithm of the U.S./Finland real exchange rate (1900-1987, annual) based on the GNP deflators as the price indices.