World-Wide Purchasing Power Parity*

Tor Jacobson and Marianne Nessén**

December 1998

Abstract

We examine long-run PPP between Germany, Great Britain, Japan and the United States over the period 1930 - 1996 using multivariate cointegration techniques. Bilateral PPP between the four countries is examined in one system (as opposed to e.g. series of trivariate systems). In all of the statistical analysis, asymptotic tests are augmented by parametric bootstrap analogues, whereby we reduce, if not eliminate, the size distortion typically present in small-sample studies. The cointegration analysis provides support for the necessary conditions for PPP (i.e. cointegrating relations are found) but not for the sufficient conditions (i.e.the coefficients in the cointegrating relations are far from what PPP predicts). These results are at odds with results from other studies that also analyze long-horizon data sets.

Key words: Long-run purchasing power parity, multivariate cointegration analysis, bootstrap inference.

JEL Classification: F30, C15, C32.

*Comments by Stefan Norrbin, Lee Ohanian, Anders Vredin, seminar-participants at Sveriges Riksbank and at the European Economic Association Annual Congress in Berlin are gratefully acknowledged.

**Research Department, Sveriges Riksbank, 103 37 Stockholm, Sweden. *E-mail*: tor.jacobson@riksbank.se and marianne.nessen@riksbank.se

1 Introduction

Does purchasing power parity hold in the long run? Are real exchange rates mean-reverting? A reading of the voluminous literature on this matter appears to give the following conclusions. If one applies unit root tests to real exchange rate data spanning long periods of time (say, close to a century or more) then evidence of long-run PPP is often found (see e.g. Frankel (1986), Abuaf and Jorion (1990) and Lothian and Taylor (1996)). However, when examining the recent post-Bretton Woods period of floating exchange rates the answer is less clear-cut. Conventional unit-root tests do not find evidence of PPP, while other approaches, e.g. using panel data, have provided evidence in favor of PPP (however, see O'Connell (1998) for a critical assessment of panel data studies).

In this paper we re-examine the case of PPP using long data sets, even though many would consider it a case closed. There are several reasons why we consider a re-examination warranted. First, earlier studies using long-run horizon data sets have typically analyzed the real exchange rate using various univariate techniques.[1] In contrast, we cast the analysis in terms of multivariate cointegration. The advantage of such a framework, as we see it, is described below. A second reason concerns size distortion, i.e. the erroneous rejection of a true null hypothesis due to an inappropriate asymptotic approximation. Engel (1996) argues that the unit-root tests referred to above may have serious size biases, leading to the false conclusion of a stationary real exchange rate. In this paper, all the asymptotic tests are complemented with parametric bootstrap analogues whereby we achieve an almost correct α-size test. Finally, in a careful re-examination of the influential paper by Abuaf and Jorion (1990) we have discovered that their conclusion concerning the validity of long-run PPP may not hold in the light of new methods and new data. We return to each of these issues shortly.

We examine annual data for the years 1930 - 1996 for Germany, Great Britain, Japan and the US, and the results of our analysis are the following. We do find evidence of cointegration between nominal exchange rates and prices; in fact the number of cointegrating vectors is exactly what PPP predicts. But the coefficients in the cointegrating vectors are far from what is compatible with PPP. Hence, we reject PPP.[2] We discuss this result in our

[1] In addition to the references cited in the text, influential papers include Diebold, Husted and Rush ((1991), Glen (1992) and Edison (1987).

[2] Some would actually interpret our results as evidence of 'weak form' PPP; see e.g.

concluding section, Section 4. Prior to that, Section 2 explains the implications of PPP in terms of cointegration. Section 3 contains the cointegration analysis.

Multivariate framework — "World-Wide PPP"

In contrast with most earlier studies of long-horizon data sets, we cast the analysis in terms of multivariate cointegration.[3] The multivariate nature of the framework offers two advantages. *First*, we are able to test for (bilateral) PPP between all countries in one system, meaning that the interdependent nature of the foreign exchange markets is taken into explicit account. Ideally, such an analysis should include prices and exchange rates of all large economies in order to fully account for the simultaneity. But doing so one would of course run into problems with degrees of freedom. Hence we have restricted the number of countries in the analysis to the four mentioned above, concentrating on what we believe to be major economies/currencies of the twentieth century. Furthermore, in this multivariate setup we will test not only individual bilateral PPP relations, but also whether all bilateral PPP relations hold simultaneously – i.e world-wide PPP. *Second*, nominal exchange rates and prices enter separately into the analysis. Hence no a priori restrictions are imposed on the joint behavior of prices and exchange rates (i.e. the so-called symmetry and proportionality conditions are not imposed, but instead subsequently tested for).[4]

MacDonald (1993). We prefer to associate PPP with the stricter requirement that the cointegrating relations satisfy certain linear restrictions. This is explained more in section 2.

[3]Earlier cointegration studies using long-horizon data sets – Kim (1990) and Ardeni and Lubian (1991) – analyze nominal exchange rates and price ratios separately using the bivariate Engle-Granger two-step procedure.

[4]Earlier studies using the multivariate cointegration setup to analyze long-run PPP — Cheung and Lai (1993), Kugler and Lenz (1993), Johansen and Juselius (1992), MacDonald (1993) and Edison, Gagnon and Melick (1997) — have used data from the post-Bretton Woods period only. Furthermore, these studies have examined PPP in series of trivariate systems (an exception is Nessén (1996)). The typical result in these studies (and Nessén (1996) is that evidence of cointegration is found, but that the cointegrating relations fail comply with the restrictions implied by PPP.

Size and power issues in tests of long-run PPP

In the empirical PPP literature there has been much concern with issues of statistical power of the tests used when examining whether real exchange rates are mean-reverting (see e.g. Cheung and Lai (1998)). On the other tack, Engel (1996) has shown that these tests may in fact have serious size biases when applied to random variables that contain a stationary but persistent component and a non-stationary component.

There is reason to believe that the usefulness of multivariate maximum likelihood cointegration analysis can be severly hampered by the curse of dimensionality, i.e. a large number of parameters in relation to a small number of observations. One undesireable effect is that the use of asymptotic critical values may jeopardize the validity of inference. This has been empirically verified in Jacobson, Vredin and Warne (1998). Gredenhoff and Jacobson (1998) have confirmed and examined the presence of size distortion for likelihood ratio tests of linear restrictions on cointegrating vectors. However, they also found that parametric bootstrap testing is a robust alternative to asymptotic approximations, even for quite large systems and as few observations as 40. In this paper, all the asymptotic tests (not only those on linear restrictions on cointegrating vectors) are augmented by parametric bootstrap analogues.

Some earlier studies

A much cited study of PPP using long-horizon data is Abuaf and Jorion (1990). We use the same dataset (although updated) and consequently we have been able to reproduce their results. Two conclusions arise: i) using more modern unit-root tests (than were available to Abuaf and Jorion) the hypothesis of a unit root in the real exchange rate is instead often **not** rejected; ii) applying the same method as Abuaf and Jorion used to an updated data set (1900 - 1996) the conclusions often change, so that a unit root in fact cannot be rejected. Another frequently cited paper is Diebold, Husted and Rush (1991). Yet, as the authors themselves point out (footnote 16, p.1263) their analysis (in short, allowing for fractional integration) cannot discriminate between *mean reversion* and *trend reversion*. In our view, PPP must be associated with mean reversion. Hence we believe that a re-examination of long-horizon data is warranted.

2 PPP and linear restrictions on prices and exchange rates

We examine long-run PPP between four large economies — Germany, Great Britain, Japan and the United States — and as mentioned in the introduction we do so in a multivariate setting. Hence, bilateral PPP between all pairs of countries can be examined simultaneously (and not only one relation at a time). The purpose of this section is to show how such a system is set up, and to identify the restrictions implied by long-run PPP.

With four countries there are, obviously, four prices and three linearly independent exchange rates. The cointegration analysis will thus be performed on seven variables simultaneously. A choice of exchange rates must be made, and we have chosen to work with dollar exchange rates.[5] Denote the natural logarithms of the DEM/USD, GBP/USD and JPY/USD nominal exchange rates by s_t^{DE}, s_t^{GB}, s_t^{JP} respectively. Further, let p_t^i be the natural logarithm of the price level in country i. The seven variables are stacked in the 7×1 vector x_t:

$$x_t = \left[s_t^{DE} \; s_t^{GB} \; s_t^{JP} \; p_t^{DE} \; p_t^{GB} \; p_t^{JP} \; p_t^{US} \right]'$$

If long-run bilateral PPP holds then the real exchange rates between all pairs of countries are stationary, or integrated of order 0, $I(0)$. In this four-country setting there are twelve real exchange rates, but only three are linearly independent. Choosing those relations involving only dollar (nominal) exchange rates, long-run bilateral PPP may be expressed as

$$q_t^i \equiv s_t^i - p_t^i + p_t^{US} \sim I(0) \qquad\qquad i = DE,\ GB,\ JP$$

where q_t^i is the real exchange rate between country i and the US. These three

[5] Concerning the choice of numeraire—if the analysis shows that bilateral PPP holds between Germany and the US, and also between Great Britain and the US, then bilateral PPP between Germany and Britain must hold also, since the real exchange rate between Germany and Great Britain is a linear combination of the real exchange rates between Germany and the US and Great Britain and the US. However, if the US-denominated real exchange rates are not stationary, then bilateral PPP between Germany and Great Britain may or may not hold.

equations can be summarized in matrix notation as:

$$\begin{bmatrix} q_t^{DE} \\ q_t^{GB} \\ q_t^{JP} \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} s_t^{DE} \\ s_t^{GB} \\ s_t^{JP} \\ p_t^{DE} \\ p_t^{GB} \\ p_t^{JP} \\ p_t^{US} \end{bmatrix} \sim I(0) \qquad (1)$$

or

$$q_t \equiv W' \, x_t \sim I(0) \qquad (2)$$

The matrix W represents the necessary and sufficient conditions implied by long-run PPP. Since the rank of W is 3, bilateral PPP — or equivalently the stationarity of bilateral real exchange rates — implies the existence of three stationary and linearly independent linear combinations of the seven series in x_t. These are the *necessary* conditions. The *sufficient* conditions are that the cointegrating vectors (if such are found) are of the same form as the columns of W.

3 The cointegration analysis

Our database contains annual observations of wholesale prices and nominal exchange rates (vs the US dollar) for Germany, Great Britain and Japan for the years 1930 - 1996.[6] The Appendix contains a fuller description of the data and sources, and also graphs of the exchange rate and wholesale price series.

The cointegration analysis performed in this paper employs methods developed by Johansen (1988, 1991). We begin by setting up the following vector error correction model (VECM):

$$\Delta x_t = \Gamma_1 \, \Delta x_{t-1} + \ldots + \Gamma_{k-1} \, \Delta x_{t-k+1} + \Pi x_{t-k} + \mu + \delta D_t + \varepsilon_t \qquad (3)$$

where x_t (defined above) and μ are column vectors with seven elements, the Γ's and Π are 7×7 matrices with coefficients, D_t is a $d \times 1$ vector with deterministic variables, δ is a $7 \times d$ matrix with coefficients, and ε_t is a

[6] Our initial intention was to examine the period 1900-1996, but we were unable to obtain a statistical model with acceptable properties.

Gaussian error term with zero mean and a covariance matrix Σ. The rank of Π is of central importance. If it has reduced rank equal to $r < 7$, then Π may be divided into two matrices α and β (i.e. $\Pi = \alpha\beta'$), where the $7 \times r$ matrix β contains the cointegrating vectors, i.e. $\beta' x_t$ is stationary.

In the subsequent sections we use this framework in the following way: First we estimate the number of cointegrating relations in a VECM of our seven-variable data set that satisfies standard specification tests. Second, we test hypotheses about the cointegration vectors. These hypotheses are summarized in equations (2) above.

3.1 Specification and mis-specification analysis

In applications of multivariate cointegration analysis, lag order determination is customarily carried out by checking the agreement between the properties of the estimated residuals in an unrestricted VAR model[7] and the assumption of independent, identically distributed normal variates. In this paper we instead examine the residuals of the VECM in equation (3) for the following reasons: (i) The mis-specification tests employed are, strictly speaking, valid for stationary processes only; (ii) The reference distributions for the specification tests are asymptotic and hence they may constitute poor approximations to the small sample distributions required for correct inference in samples of our size. In order to circumvent this problem we evaluate the residuals using bootstrapped versions of the mis-specification tests. But the bootstrap procedure also requires stationarity to be valid, and is thus better suited for the VECM model rather than a VAR model; (iii) Ultimately, what should be of concern is the model which we interpret, i.e. the stationary VECM model and not the non-stationary VAR model.

Tables 1a-b present mis-specification analyses for the VECM in equations (3) for $k \in \{2, 3, 4\}$ and $r \in \{1, 2, 3, 4, 5\}$ with respect to multivariate serial correlation, multivariate normality, and multivariate autoregressive conditional heteroscedasticity.[8] Inference is given in terms of asymptotic as well as bootstrapped p-values.[9] Apart from providing robust inference, the boot-

[7]That is, in a model where the reduced rank - or cointegrating - restriction has not been imposed.

[8]In all estimations, D_t contains intervention dummy variables for the years 1938, 1940, 1946 and 1951.

[9]The underlying idea in bootstrap hypothesis testing is to estimate a reference distribution (critical values). This is done by first generating a large number of pseudo samples

strap tests provide further interesting information — namely an evaluation of the size property of the corresponding asymptotic test. For instance, we see that the two sets of p-values for the multivariate ARCH test (Table 1a) often are of same order of magnitude. This is also true for the first order serial correlation test LM(1). Nonetheless, the general impression when looking at both Table 1a and Table 1b is that asymptotic mis-specification testing provides extremely poor inference.

Turning to the conclusions provided by the bootstrapped p-values, we find evidence of mis-specification for the second-order models, i.e. $k = 2$. In particular, the LM(1)-test for first order serial correlation as well as the three normality tests (Table 1b) often suggest rejection of the nulls. Likewise, the models with four lags display signs of first-order serial correlation. Hence, a parsimonious approach suggests that three lags is the appropriate choice. However, we will keep the lag order specification open and study the cointegration rank tests for all three lag orders.

[Tables 1a-b about here.]

3.2 Cointegration tests

The cointegration test results are reported in Table 2, and again we have estimated small sample reference distributions by bootstrapping in order to avoid the problems using asymptotic inference. The first panel of Table 2 shows asymptotic critical values. Using these, we would infer a cointegrating rank of $r = 4$ for a lag order of $k = 2$. For $k = 3$ the rank is also 4 for a 10% level test, and, finally, the asymptotic inference suggests three cointegrating vectors in the case of $k = 4$.[10]

with the null hypothesis in question imposed. The next step is to evaluate the test function in each pseudo sample and then arrange them in ascending order. We generate the pseudo samples by a parametric procedure where we substitute the parameters of (3) with the estimates from the original sample and feed in pseudo random vectors, generated as $N_p\left(0,\widehat{\Sigma}\right)$, in place of ε_t. All bootstrap results in this paper involve 10,000 generated pseudo samples.

[10]The tests are carried out in the following way using the $k = 3$ case as an illustration: the likelihood ratio test statistic for the hypothesis $H_0 : r = 0$ is 189.85, and comparing this with the critical values on the first row of the first panel we reject at every sensible test level. Moving down one step we consider $H_0 : r \leq 1$ and the statistic 132.22, which is also rejected, and so is $H_0 : r \leq 2$ with statistic 85.32, and $H_0 : r \leq 3$ with statistic 44.42. However, $H_0 : r \leq 4$ cannot be rejected since the test statistic 19.47 is smaller than the

[Table 2 about here.]

Turning to the remaining three panels, with bootstrapped small sample critical values, the picture changes. Comparing the three sets of critical values, we find that they increase with lag order k. An interpretation of this is that if too many lags are included then the use of asymptotic criticial values will imply a risk of inferring too many cointegrating vectors. For example, consider the $k = 3$ case. Asymptotic critical values suggest $r = 4$, while the bootstrap tests give $r = 3$. In the $k = 4$ case, the bootstrap tests find only one cointegrating vector, while the use of asymptotic critical values indicate three. The critical values in the second panel ($k = 2$) are very close to the asymptotic ones, and, of course, the two tests agree on $r = 4$. It is tempting to interpret this as evidence of $k = 2$ being an appropriate specification. However, we believe that the lack of disagreement between the asymptotic and the small-sample distributions is a sign of mis-specification of the $k = 2$ models, i.e. effects from nuisance parameters on the bootstrapped critical values cancel.

To sum up: the empirical analysis suggests a parsimonious model with three lags. Furthermore, we find evidence of three cointegrating vectors, which is compatible with the necessary conditions for PPP. The estimated β matrix is reported in Table 3. The table also contains the matrix W, derived in Section 2, which summarizes the restrictions on x_t implied by PPP. Note that $\widehat{\beta}$ has been normalized so as to make $\widehat{\beta}$ and W as similar as possible. A casual comparison between the two sets of vectors suggests that PPP does not hold – the coefficients in $\widehat{\beta}$ are very much different from the corresponding elements in W. In the next section, we will test formally for the agreement between $\widehat{\beta}$ and W.

80th percentile 23.72.

Table 3. PPP: theoretical and estimated cointegrating vectors.

$$W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \widehat{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -407.60 & -111.15 & -498.41 \\ 94.90 & 25.92 & 115.82 \\ 20.24 & 5.52 & 23.96 \\ 96.22 & 25.60 & 118.48 \end{bmatrix}$$

3.3 Testing linear restrictions

Having found support for the necessary condition for PPP, we now turn to the sufficient conditions. The multivariate setup used in this paper actually enables us to test for PPP in different ways. First, we test whether *all three* bilateral PPP relations hold – the 'world-wide' PPP. After that we examine two relations at a time. Finally, we test each relation separately.

Table 4 below reports the results from these tests, which are likelihood ratio tests of the linear restrictions on the cointegrating vectors implied by all three, different combinations of two, and individual vectors of the hypothesized vectors in equation (1), respectively. Both asymptotic and bootstrapped critical values are provided. A comparison between these two sets of critical values demonstrates that the likelihood ratio tests are indeed greatly oversized for this particular application. Nonetheless, the asymptotic test conclusions do not change when using the bootstrap test — in each case the null is rejected by a large margin. Hence, the hypothesis that all three bilateral real exchange rates are stationary is rejected, and so are the hypotheses concerning all subsequent combinations of two bilateral real exchange rates, and individual bilateral real exchange rates. This means that the rejection of 'world-wide' PPP is not due to any one particular relation, but must be attributed to the non-stationarity of all bilateral real exchange rates.

[Table 4 about here.]

In summary, we have found support for our hypothesis that the variables in x_t can be characterized by an error correction model like equation (3).

This implies that they are driven by a limited number of common stochastic trends and therefore are tied together in the long run. There are three long-run, cointegrating, relations. However, none of these long-run relations can be interpreted in terms of PPP.

4 Discussion

Previous studies of long-run purchasing power parity analyzing long-horizon data sets have predominantly used univariate techniques (e.g. unit-root tests) and have often found support for long-run PPP. We, on the other hand, use a multivariate approach, and arrive at a different conclusion. We do find cointegrating vectors between nominal exchange rates and prices - and just the number that PPP would predict - but none of these can be interpreted in terms of PPP.

It is difficult to reconcile the evidence given by traditional unit-root tests with the results provided in this study. What can explain this striking difference in conclusion? One possible explanation is offered by Engel (1996), who argues that the traditional unit-root tests are greatly over-sized. The reliability of our results is enhanced by what we believe to be a well-specified statistical model and by the fact that all the asymptotic tests have been replaced by robust bootstrap inference.

Now, whereas the bootstrap test can be expected to be approximately correct in size, it should be noted that its power will not be higher, nor lower, than the power of the size-adjusted asymptotic test. This has been theoretically predicted for the general case by Davidson and McKinnon (1996) and verified for the likelihood ratio test of linear restrictions on cointegrating vectors by Gredenhoff and Jacobson (1998) using Monte Carlo simulation. Moreover, the results in Gredenhoff and Jacobson suggest that the power of the likelihood ratio test in a complex model based on relatively few observations, such as the one at hand, cannot be expected to be high. Despite this we do reject the null of PPP. Had we not, low test power could very well have driven that result. In other words, the bootstrap procedure ensures a proper size for the test and the insufficient power only strengthens the rejection result.

The conclusion arising from our analysis is that real exchange rates are non-stationary, even when examining data streching over long periods of time. Hence shocks to real exchange rates do not subside with time, but instead have infinately long-lived effects. This suggests that permanent real shocks are the predominant source of real exchange rate movements. A natural suggestion for future research is thus to develop models of real exchange rate behavior that focus mainly on real factors.

Appendix

The database is comprised of three nominal exchange rates and four wholesale price indices. The frequency is annual and the series run from 1930 to 1996. See graphs A1 and A2. For the period up to 1972, the source is Lee (1976) after which data has been obtained from the IMF's International Financial Statistics (CD-Rom).The exchange rates are the price of USD in DEM, GBP, and JPY respectively. The WPI's are from row 63 in the IFS-tapes. In the Lee-data there are six missing observations in the DEM/USD exchange rate (1942-47) and four in the WPI (1944-47). These have been replaced by linear intrapolation.

[Graphs A1 and A2 about here.]

References

[1] Abuaf, N. and P. Jorion (1990) " Purchasing Power Parity in the Long Run" *Journal of Finance*, vol. 45.

[2] Ardeni, P. and D. Lubian (1991), "Is there trend reversion in purchasing power parity?", *European Economic Review*, vol. 35.

[3] Cheung, Y. and K. Lai (1993), "Long-Run Purchasing Power Parity During the Recent Float", *Journal of International Economics*, vol. 34.

[4] Cheung, Y. and K. Lai (1994), "Mean reversion in real exchange rates", *Economics Letters*, vol 46.

[5] Cheung, Y. and K. Lai (1998), "Parity reversion in real exchange rates during the post-Bretton Woods period", *Journal of International Money and Finance*, 17.

[6] Davidson, R. and J.G.McKinnon (1996), "The Power of Bootstrap Tests", Queen's Institute for Economic Research, Discussion Paper no. 937.

[7] Diebold, F., S. Husted and M. Rush (1991), "Real Exchange Rates under the Gold Standard", *Journal of Political Economy*, vol 99.

[8] Doornik, J. A. and H. Hansen (1994), "A Practical Test for Univariate and Multivariate Normality," manuscript, Nuffield College.

[9] Edison, H. (1987), "PPP in the Long Run: A Test of the Dollar/Pound Exchange Rate (1890 - 1978), *Journal of Money, Credit and Banking*, vol. 19, no. 3.

[10] Edison, H., J. Gagnon and W. Melick (1997) "Understanding the empirical literature on purchasing power parity: the post-Bretton Woods era", *Journal of International Money and Finance*, vol 16, no.1.

[11] Engel, C. (1996), "Long-Run PPP may not hold after all", NBER working paper 5646.

[12] Frankel, J. (1986), "International Capital Mobility and Crowding-Out in the US Economy: Imperfect Integration of Financial Markets of or Goods Markets?" In: Hafer, R. (Ed.), How Open is the US Economy? Lexington Books.

[13] Glen, J. (1992), "Real exchange rates in the short, medium, and long run", *Journal of International Economics*, vol. 33, pp. 147-166.

[14] Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press.

[15] Gredenhoff, M. and T. Jacobson (1998), "Bootstrap Testing and Approximate Finite Sample Distributions for Tests of Linear Restrictions on Cointegrating Vectors", Sveriges Riksbank Working Paper Series no. 67, April.

[16] Jacobson, T., A. Warne and A. Vredin (1998), "Are Real Wages and Unemployment Related?", *Economica*, vol. 65, 69 - 96.

[17] Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control,* vol. 12, pp. 231-254.

[18] Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica,* vol. 59, no. 6.

[19] Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

[20] Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration — With Applications to the Demand For Money." *Oxford Bulletin of Economics and Statistics,* vol. 52, no. 2.

[21] Johansen, S. and K. Juselius (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK." *Journal of Econometrics* 53, 211-244.

[22] Kim, Y. (1990), "PPP in the Long Run: A Cointegration Approach." *Journal of Money, Credit and Banking,* vol. 22, no. 4.

[23] Kugler, P. and C. Lenz (1993), "Multivariate Cointegration Analysis and the Long-Run Validity of PPP", *Review of Economics and Statistics*, vol. 75.

[24] Lee, M.H. (1976), *Purchasing Power Parity*, Marcel Dekker Inc., NY

[25] Lothian, J. and M. Taylor (1996) "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries" *Journal of Political Economy*, vol 104, no 3, June.

[26] MacDonald, R. (1993), "Long-Run Purchasing Power Parity: Is It For Real?", *Review of Economics and Statistics*, vol. 75.

[27] Mardia, K.V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications", *Biometrika*, 57, 519-30.

[28] McLeod, A.I. and W.K. Li (1983),"Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations", *Journal of Time Series Analysis*, 4, 269 - 73.

[29] Nessén, M. (1996) "Common Trends in Prices and Exchange Rates. Tests of Long-Run Purchasing Power Parity", *Empirical Economics*, vol. 21, pp. 381 - 400.

[30] O'Connell, P. (1998), "The Overvaluation of Purchasing Power Parity", *Journal of International Economics*, vol 44, pp.1-19.

Table 1a. Multivariate specification tests with respect to serial correlation and autoregressive conditional heteroscedasticity

Model	*Portman*	p-value^a	p-value^b	*LM(1)*	p-value^a	p-value^b	*M-ARCH*	p-value^a	p-value^b
$k=2, r=1$	827.4345	.0060	.8541	79.9893	.0034	.0055	766.2913	.6678	.9206
$r=2$	835.8984	.0019	.9368	79.3850	.0039	.0093	736.8913	.8844	.9830
$r=3$	831.1548	.0015	.9868	72.2412	.0170	.0482	760.1752	.7228	.9881
$r=4$	819.2090	.0021	.9976	65.9406	.0535	.1367	785.5800	.4774	.9817
$r=5$	820.8309	.0010	.9994	62.5445	.0926	.2198	792.3770	.4100	.9891
$k=3, r=1$	768.3354	.0096	.9999	58.1933	.1729	.2141	730.7170	.9132	.9996
$r=2$	787.4386	.0013	.9999	50.5119	.4136	.5051	720.7186	.9481	.9999
$r=3$	766.1343	.0039	.9999	54.4689	.2743	.4055	740.9107	.8626	.9999
$r=4$	775.2380	.0011	.9999	53.1104	.3188	.4956	758.0056	.7412	.9999
$r=5$	769.2176	.0009	.9999	59.4987	.1447	.2815	775.9844	.5739	.9998
$k=4, r=1$	775.8646	.0000	.9999	64.1266	.0721	.0904	856.6166	.0012	.9987
$r=2$	790.4096	.0000	.9999	75.8234	.0083	.0164	870.8842	.0004	.9998
$r=3$	807.0496	.0000	.9999	73.3785	.0136	.0535	880.7524	.0002	.9999
$r=4$	812.4829	.0000	.9999	78.4268	.0048	.0304	928.4012	.0000	.9989
$r=5$	813.9777	.0000	.9999	77.4685	.0059	.0538	920.2695	.0000	.9999

Remarks: *Portman* is the multivariate Portmanteau test, which is asymptotically χ^2 with degrees of freedom $n^2([T/4]-k+1)-nr$. *LM(1)* is a Lagrange multiplier test for first-order serial correlation, which is asymptotically χ^2 with n^2 degrees of freedom. *M-ARCH* is a multivariate version of the univariate Lagrange multiplier test against autoregressive conditional heteroscedasticity suggested by Granger and Teräsvirta (1993) (cf McLeod and Li, 1983). Since an asymptotic distribution remains to be derived, we have used critical values from a χ^2 with $n^2([T/4])$ degrees of freedom for the asymtotic inference. p-value^a refers to the p-value given by use of asymptotic critical values, and p-value^b is the bootstrap analogue.

Table 1b. Multivariate specification tests with respect to normality

Model	*Omnibus*	p-value^a	p-value^b	*Skew*	p-value^a	p-value^b	*Kurt*	p-value^a	p-value^b
$k=2, r=1$	64.4028	.0000	.0005	175.0616	.0000	.0019	120.0585	.0000	.0000
$r=2$	50.2674	.0000	.0084	174.9861	.0000	.0034	88.7415	.0000	.0013
$r=3$	39.0316	.0004	.0616	144.4784	.0000	.0771	60.0070	.0000	.0501
$r=4$	37.2622	.0007	.0988	126.0577	.0021	.2739	47.2204	.0000	.1985
$r=5$	30.9683	.0056	.2429	116.5163	.0110	.4889	36.8619	.0000	.5122
$k=3, r=1$	61.8237	.0000	.0281	169.3063	.0000	.1234	93.0652	.0000	.1502
$r=2$	73.3462	.0000	.0093	167.7434	.0000	.1577	83.5964	.0000	.2991
$r=3$	70.0157	.0000	.0187	158.4301	.0000	.2824	60.9098	.0000	.7537
$r=4$	61.6742	.0000	.0618	134.0610	.0004	.6924	47.7993	.0000	.9643
$r=5$	51.3362	.0000	.1982	126.5571	.0019	.8220	37.9216	.0000	.9977
$k=4, r=1$	71.9059	.0000	.0676	195.7940	.0000	.1781	127.6232	.0000	.3895
$r=2$	63.3797	.0000	.2753	163.7948	.0000	.6984	98.2109	.0000	.9503
$r=3$	69.6121	.0000	.1844	176.5531	.0000	.5468	102.1468	.0000	.9111
$r=4$	53.3646	.0000	.5059	144.0468	.0001	.8996	65.9231	.0000	.9996
$r=5$	47.1243	.0000	.7199	106.0071	.0527	.9989	43.4372	.0000	.9999

Remarks: *Omnibus* refers to the multivariate test for normality suggested by Doornik and Hansen (1994), which is asymptotically χ^2 with $2n$ degrees of freedom. *Skew* and *Kurt* are Mardia's (1970) multivariate tests for excess skewness and kurtosis. They are both asymptotically χ^2 with $\frac{n}{6}(n+1)(n+2)$ d.f. and 1 d.f., respectively. p-value^a refers to the p-value given by use of asymptotic critical values, and p-value^b is the bootstrap analogue.

Table 2. Asymptotic and bootstrapped distributions for the LR cointegration test

Asymptotic distribution for an unrestricted constant							*trace statistic*		
r	$p-r$	80%	90%	95%	97.5%	99%	*k = 2*	*k = 3*	*k = 4*
0	7	111.79	117.73	123.04	127.59	133.04	257.60	189.85	187.83
1	6	84.10	89.37	93.92	97.97	102.95	165.62	132.22	125.65
2	5	60.23	64.74	68.68	72.21	76.37	100.57	85.32	79.89
3	4	40.08	43.84	47.21	50.19	53.91	55.53	44.42	42.49
4	3	23.72	26.70	29.38	31.76	34.87	21.98	19.47	17.94
5	2	11.06	13.31	15.34	17.24	19.69	5.38	2.94	3.66
6	1	1.64	2.71	3.84	5.02	6.64	0.00	0.19	1.13

Bootstrapped distribution for the empirical VAR model, k = 2							
r	$p-r$	80%	90%	95%	97.5%	99%	*trace statistic*
0	7	109.05	117.47	124.00	129.76	138.01	257.60
1	6	77.31	84.37	90.61	95.97	101.84	165.62
2	5	50.88	56.70	61.95	66.51	72.20	100.57
3	4	31.55	36.40	40.18	44.51	48.96	55.53
4	3	19.50	23.33	26.64	29.75	34.00	21.98
5	2	9.00	11.55	14.10	16.75	20.10	5.38
6	1	2.36	3.89	5.38	7.04	9.62	0.00

Bootstrapped distribution for the empirical VAR model, k = 3							
r	$p-r$	80%	90%	95%	97.5%	99%	*trace statistic*
0	7	134.96	144.86	153.14	161.47	170.65	189.85
1	6	104.69	113.10	120.98	127.71	136.77	132.22
2	5	74.77	82.02	88.18	94.02	101.86	85.32
3	4	46.13	52.03	57.71	62.66	67.93	44.42
4	3	30.73	35.42	39.79	44.18	49.73	19.47
5	2	10.85	13.94	16.94	19.45	23.02	2.94
6	1	2.72	4.47	6.35	8.58	11.11	0.19

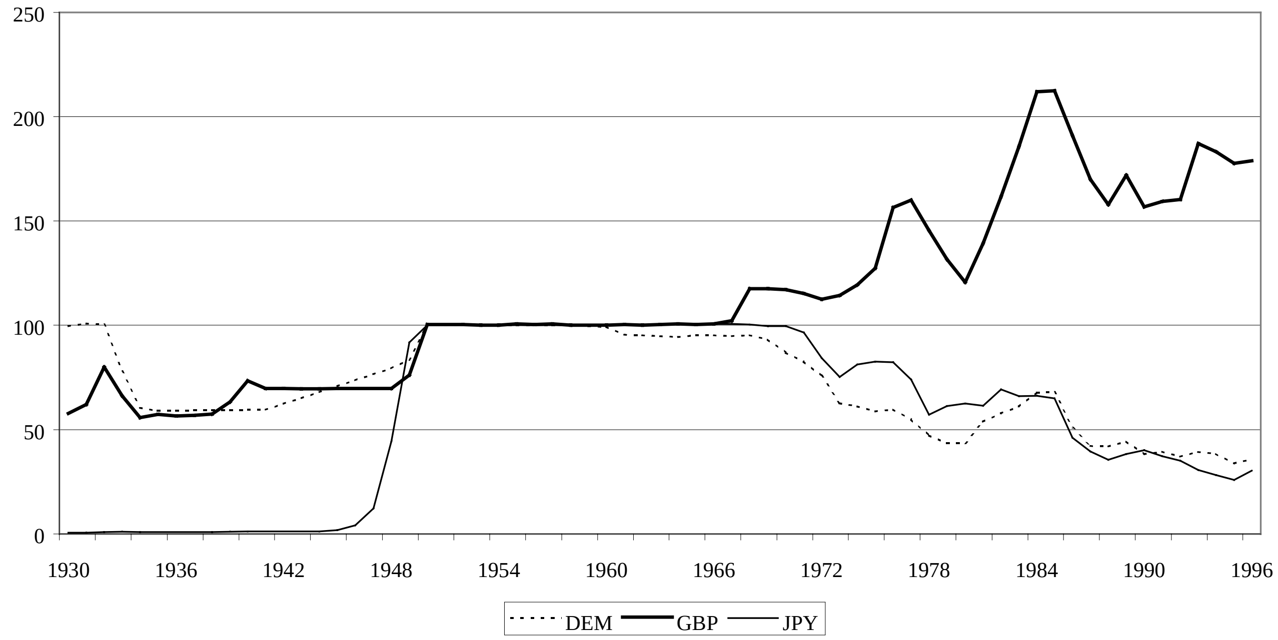
Bootstrapped distribution for the empirical VAR model, k = 4							
r	$p-r$	80%	90%	95%	97.5%	99%	*trace statistic*
0	7	166.17	178.61	188.16	197.95	208.85	187.83
1	6	129.50	141.13	151.90	160.05	170.21	125.65
2	5	119.50	129.64	138.77	146.90	156.39	79.89
3	4	74.02	81.26	87.79	93.13	99.45	42.49
4	3	35.40	41.60	47.02	52.95	59.28	17.94
5	2	14.27	18.62	22.76	27.17	31.61	3.66
6	1	3.77	6.38	8.67	11.23	14.50	1.13

Remarks: The first set of asymptotic critical values are from Table 15.3 in Johansen (1995). The empirical distribution are estimated using a parametric bootstrap procedure.

Table 4. Testing linear restrictions: asymptotic and bootstrapped percentiles

Hypothesis	*LR test statistic*	*Asymptotic percentiles*				*Bootstrapped percentiles*		
		90%	*95%*	*99%*		*90%*	*95%*	*99%*
$q_t^{DE}, q_t^{GB}, q_t^{JP} \sim I(0)$	86.08	18.55	21.03	26.22	$\chi^2\,(12)$	39.49	44.89	55.84
$q_t^{DE}, q_t^{GB} \sim I(0)$	57.18	13.36	15.51	20.09	$\chi^2\,(8)$	27.87	32.12	42.09
$q_t^{DE}, q_t^{JP} \sim I(0)$	54.70				$\chi^2\,(8)$	27.83	32.33	41.54
$q_t^{GB}, q_t^{JP} \sim I(0)$	54.59				$\chi^2\,(8)$	28.06	31.98	41.28
$q_t^{DE} \sim I(0)$	31.87	7.78	9.49	13.28	$\chi^2\,(4)$	16.29	19.97	28.28
$q_t^{GB} \sim I(0)$	36.45				$\chi^2\,(4)$	17.38	20.97	29.48
$q_t^{JP} \sim I(0)$	33.69				$\chi^2\,(4)$	15.49	18.88	26.12

Graph A1: Nominal exchange rates vs USD
Index 1958 = 100



Graph A2: Wholesale Price Indices
Index 1958 = 100